

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:10:54 ; Search time 30 Seconds

(without alignments)
949.174 Million cell updates/sec

Title: US-09-989-981A-8

Sequence: 1 MAGKAEERGLPKGATPDT.....FMVLYVSLRFIKRPSODW 673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/CTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Dackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query length	DB ID	Description
1	640.5	18.3	655	4	US-09-245-808-1
2	638.5	18.2	655	4	US-09-767-594-1
3	250	7.1	1684	3	US-08-665-259-25
4	250	7.1	1684	3	US-08-762-500-25
5	250	7.1	1704	3	US-08-762-500-75
6	243	6.9	593	4	US-09-252-991A-21665
7	242	6.9	330	4	US-09-252-991A-21665
8	230	6.6	370	4	US-09-252-991A-20719
9	228.5	6.5	359	4	US-09-328-352-6329
10	227	6.5	432	4	US-09-328-352-6798
11	225	6.4	345	4	US-09-252-991A-18872
12	222.5	6.3	1457	3	US-08-665-259-27
13	222.5	6.3	1457	3	US-08-762-500-27
14	219.5	6.3	396	4	US-08-311-731A-209
15	219.5	6.3	676	4	US-09-252-991A-21751
16	217.5	6.2	235	4	US-09-198-452A-242
17	217	6.2	683	4	US-09-252-991A-26621
18	216.5	6.2	345	4	US-09-107-532A-3849
19	215	6.1	250	4	US-09-328-352-7153
20	215	6.1	256	4	US-09-134-001C-4600
21	214.5	6.1	376	2	US-08-997-080-89
22	214.5	6.1	376	2	US-08-997-362-89
23	214.5	6.1	376	3	US-08-873-970-89
24	214.5	6.1	376	3	US-09-095-855-89
25	214.5	6.1	376	4	US-09-324-542-89
26	214.5	6.1	376	4	US-09-205-426-89
27	213.5	6.1	383	4	US-09-252-991A-25307

28	212	6.0	707	3	US-08-772-270A-4	Sequence 4, Appl
29	212	6.0	707	4	US-09-062-126-4	Sequence 4, Appl
30	211.5	6.0	711	3	US-08-772-270A-12	Sequence 12, Appl
31	210	6.0	929	4	US-09-252-991A-22946	Sequence 22946, A
32	209.5	6.0	217	4	US-09-107-532A-6401	Sequence 6401, Ap
33	209.5	6.0	402	4	US-09-107-532A-5360	Sequence 5360, Ap
34	209.5	6.0	406	4	US-09-134-001C-3570	Sequence 3570, Ap
35	209.5	6.0	460	4	US-09-134-001C-3369	Sequence 3369, Ap
36	208	5.9	256	4	US-09-107-532A-4208	Sequence 4208, Ap
37	208	5.9	260	4	US-09-107-532A-5350	Sequence 5350, Ap
38	207.5	5.9	261	4	US-09-252-991A-33060	Sequence 33060, A
39	207.5	5.9	655	4	US-09-634-238-289	Sequence 289, App
40	207	5.9	227	4	US-09-134-001C-5550	Sequence 5550, Ap
41	207	5.9	337	4	US-09-328-352-6905	Sequence 6905, Ap
42	207	5.9	369	4	US-08-463-092B-7	Sequence 7, Appl
43	205.5	5.9	1548	1	US-08-460-907B-7	Sequence 21625, A
44	205.5	5.9	1548	2	US-09-252-991A-21625	
45	204	5.8	624	4	US-09-252-991A-21625	

ALIGNMENTS

```
RESULT 1
US-09-245-808-1
; Sequence 1, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; FILE REFERENCE: encodes it
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Human MCF-7/Adr-tp cells
US-09-245-808-1

Query Match      18.3%; Score 640.5; DB 4; Length 655;
Best Local Similarity 27.2%; Pred. No. 3.1e-62;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

QY      19 DTSGIADRLFSSESNSLYFTYSGQPNILEVDNLQVDLASOVFEQLAQFKMPWSP 78
      16 NTNG-----FPATASNDLKAFTEGA--VLSFNICYRVALKSGF-----LP----- 54
      79 SCQNCSEIGI-ONLSEFKVSGOMLAIIGSSGGRASRNOVITGRGCGIKSGQIWINQ 137
      55 -CRKVEKEITLSNIGIMKPG-LNALIGTGGKSSLDVLVLAARDPSGL-SGDVLINCA 111
DB      138 PSSPOLVRKC-VAHYRHOHNLPLNLTRETLEAFIAQMRLPFFSOAQRKREVDYAEIR 196
      112 PRPNPF--KCNNGYVADVDDVWCVGLTRVRLNLOFSAAKRLATMTVHKKRIRNRYIOELG 169
QY      197 LRQCDTRVGNMNVYGLSGGERRRVSIGVOLLMPNGIILDEPTSGLDSTFNAHLVKTL 256
      170 LDKVADSKYGTQFINGVSGGERRKRTSIGMELTIDPSILFIDPTGLDSTFNAVLILK 229
DB      257 RLAKNRLVLSLHOPRSDIFRLFDVLVLTMTSGTPIYLGAQHMYQYFAIGPPRRYSN 316
      230 RNSKQGRITFESIHOPRYSIFRLFDSTLLASGLRMLFHPQADQALGYFSAGYHCEAVNN 289
QY      317 PADFVVDYDLSIDR-----SREQLATVRE--KAQSLAALFLKVRDL--DDFLKKAETK-- 366
      290 PADFLDIITINGSTAVALNRREDPKATLIEESKQDKPDLIEKLAELIYVNSFYK-ETKAE 348
```

QY 367 -----DDEDTCVESSTYPLDTNCLSPTRKMPGAVOQFTLLIRROISNDRDLPPTLLI 419
DB 349 LHOISGGEKKKITYEKEISYTTSTFC-----HOLRWVSKRSFKNLGNQASIA 397
QY 420 HGAECALMSMTIGFYEGHGSIOLEPMDTALLFMGALIPRNVLIDVSKISY----- 473
DB 398 OIIVTVLGLVIGALFYEGHGSIOLEPMDTALLFMGALIPRNVLIDVSKISY----- 473
QY 474 -----ERAMLYELEDGLYTTGPFYFAKILGE-LPEHCAYIIYGMPTWLANLRDQLPFL 528
DB 447 LEVVEKLEFIHEIYISGYRVSSYFLGKILSDLPMLPSIITCTIVFMGLKPKADAF 506
QY 529 LHFLVLMVVECCRIMALAAALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPA 588
DB 507 FVMEFTLMVAVASASMAIAAGSVSVATLMTICFVFMIFSGLLVNLTTIASWLS 566
QY 589 MISVSEFLMCEGEMKIOFSRRTYKMLGNLT-----IAGSDKIL--SAMELDSYPL 639
DB 567 WLOYSIPRYGFTALQHNEFLGQNF-CPGLNATGNPCNYATCTGEYLVKQIDLSPMWG 625
QY 640 LYAIVLYIGLSGFMVLYVSLRPIKQ 667
DB 626 LMKHNVALACMIVITLTAIVLKLFLPK 653

RESULT 2
US-09-767-594-1
Sequence 1, Application US/09767594
Patent No. 6521635

GENERAL INFORMATION:
APPLICANT: Bates, Susan
APPLICANT: Robey, Robert
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
FILE REFERENCE: 015280-402100US
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: US 60/177,410
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
US-09-767-594-1

Query Match 18.28; Score 638.5; DB 4; Length 655;
Best Local Similarity 27.98; Pred. No. 5.2e-62;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

QY 80 CONSCELGI-QNLSFKVRSOMLAIIGSGGRASLDVITGRGHGKIKSGOIMWIGOP 138
DB 55 CRKVEKEILSNINIGIKRP-LNAILPTGGKSLDVLAAKRDPSGL-SGDVLINGAP 112
QY 139 SSPOLVRKC-VAHROHNOQLPNTVRETLAFLAOMKLPPTFSQORDKREVEDVIAELRL 197
DB 113 RPAHF--KCSNGYVQDDVWGLTVRENLOFSMALRLATVTMTHKHEKNEINRYIEBGL 170
QY 198 ROCADTRVGMNVYVGLSGGERRRSTIVOLAMPNGIILDEPPISGLDSFAHNLVYTLR 257
DB 171 DKVADSVYGOFTINGVSGGKRRKISIMELITDPSIILDEPTGLDSSSTANAVLLKLKR 230
QY 258 LANGNRLLVLSLHOPRSDFLEFDVLMTSGTPIYGAOHMVQYFTALGYDQPRYSNP 317
DB 231 MSKGRITIFSIHQPRRSIFKRDPSLTLASGRMLFPGRAOENALGYESAGYCEAYNMP 290

QY 318 ADFVYDLTSDIR-----SREOLATRE--KAOSLAALFEKVRDL--DOFLKAEK--- 366
DB 291 ADFVYDLTSDIR-----SREOLATRE--KAOSLAALFEKVRDL--DOFLKAEK--- 366
QY 367 -----DDEDTCVESSTYPLDTNCLSPTRKMPGAVOQFTLLIRROISNDRDLPPTLLI 420
DB 398 OIIVTVLGLVIGALFYEGHGSIOLEPMDTALLFMGALIPRNVLIDVSKISY----- 473
QY 421 HGAECALMSMTIGFYEGHGSIOLEPMDTALLFMGALIPRNVLIDVSKISY----- 473
DB 399 IIVTVLGLVIGALFYEGHGSIOLEPMDTALLFMGALIPRNVLIDVSKISY----- 473
QY 474 -----ERAMLYELEDGLYTTGPFYFAKILGE-LPEHCAYIIYGMPTWLANLRDQLPFL 529
DB 448 FVEVEKLEFIHEIYISGYRVSSYFLGKILSDLPMLPSIITCTIVFMGLKPKADAF 507
QY 530 LHFLVLMVVECCRIMALAAALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPA 589
DB 508 FVMEFTLMVAVASASMAIAAGSVSVATLMTICFVFMIFSGLLVNLTTIASWLS 567
QY 590 ISKVSFLMCEGEMKIOFSRRTYKMLGNLT-----IAGSDKIL--SAMELDSYPL 640
DB 568 LQYSIPRYGFTALQHNEFLGQNF-CPGLNATGNPCNYATCTGEYLVKQIDLSPMWG 626
QY 641 LYAIVLYIGLSGFMVLYVSLRPIKQ 667
DB 627 LMKHNVALACMIVITLTAIVLKLFLPK 653

RESULT 3
US-08-665-259-25
Sequence 25, Application US/08665259
Patent No. 6028173

GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1684 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-665-259-25


```

TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: 1linear
MOLECULE TYPE: protein
US-08-762-500-75

Query Match
Best Local Similarity 7.1%; Score 250; DB 3; Length 1704;
Matches 124; Conservative 85; Mismatches 176; Indels 116; Gaps 27:

QY 82 NSCEIGIONLSFKVSGOMLAIIGSGGRASLDVITGRGKGKISQGIWINGOPSSP 141
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 543 NDRRAVRDLNLTNLGEGITVLGNHAGKTTLSMLTGL--PPTSGRAYISGEISQ 599
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 QLV--KCVAHYRONHOLLPNLTVRETLAFIQMLPRFSAQDKREVDYIAELRLQ 199
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 600 DMVQIRKSLGLCPQHDILFDNLTVAEHLFYAQLK-----GLSRQKPEEYKQMLHIIG 653
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 CADTFVGMNYVRLSGGERRRVSIGVOLLMPGILLDEPTSGIDSEFTAHNLVKTLSRLA 259
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 654 LEDK--WNSRSRFLSGMRKRKISGIALAGSKVILDEPTSGMAISRAIWDLOR-Q 710
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 KGNRLVLSLH-QPRSDIFRFLDVLVMTS-----GTPITYL-----GAQHMYOYFTAG 308
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 711 KSDRIIVLTTHPEADL--LGDRIAYMAKGLQCCSSFLFKQYGYGHR-----TLVK 764
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 YCPRYSNPADF-----YVDLTISDRSREOLA--TREKQSLALFL--EKVRL 356
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 765 EP--HCPNEDISQLVHHHPNATE--SSAGALSFTLPRSTHREGEIFALEKQKEL 820
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 357 DDFLKAETKDLDE-----DTCVESSVT-----PLDTNCLPS--P 389
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 821 GIASGASITMEVEFLVNGKLYDSSMDIOATQLPALQYHRRASDAVDSNLCGADP 880
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 390 TKMPAV-----QOFTTLIRQISDFRDLPTLLIHGAEAC--LMSM 429
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 881 SDGIALIEERTAVAKLTGLALHCOQFAMFLKKAAYSWREKMY--AAQVLPVLCV 937
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 TIGLFGHGSITQSLFMDTALLFMIG--ALIFPNVI-LDVISKCTSERMAYLEED 484
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 938 TLALLAINYS---ELFDDPMRLTLGEGRTVVPFVSGTSQLCOQLSE-----HLMD 988
QY 485 GLYTG--PYFFAKIGELPE 503
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 989 ALQAEQGE--REVGLDEE 1006

RESULT 6
US-09-252-991A-21665
Sequence 21665, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21665
LENGTH: 593
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21665

Query Match
Best Local Similarity 6.9%; Score 243; DB 4; Length 593;
Matches 74; Conservative 61; Mismatches 100; Indels 34; Gaps 9;

QY 89 QNLSFKVSGOMLAIIGSGGRASLDVITGRGKGKISQGIWINGO---PSSPOLVR 145
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 282 RQIDPAAAGFEVTLTGSGCGKSTLNCIAL--TEVDSGRILIDEDVPLPPO--K 336
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 KCVAHYRONHOLLPNLTVRETLAF--IAQMLPRFSAQDKREVDYIAELRLQCACTR 204
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 337 RQIANVFOSTALFPMTYQONAVAFGLRMQKVP---AAELKORVAELTELGLGYA--- 389
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 VGNMYVRLSGGERRRVSIGVOLLMPGILLDEPTSGIDSEFTAHNLVKTLSRLAKGNRL 264
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 390 --ARYPHQLSGGOCQVALRSLVTRPRILLDEPLSLMDARIRKHLRQRIQOQELGL 447
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 VLSHQRSDIFRFLDVLVMTSGTPITYLGAQHMYOYFTAGIYPCPRYSNPADF--V 322
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 448 TTVFTYHDOEALTLSDRIIVLMNAGRIVQSGDAETL-----YTAPENAFAGFIQNY 499
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 323 DLTISDRSR-----EOELATREKAQSL 345
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 500 NLLDPAQSRLLGQPCAOQVAILRPSLRL 528

RESULT 7
US-09-252-991A-27569
Sequence 27569, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON.
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27569
LENGTH: 330
TYPE: PRT

```


Db 27 LDKGLIMQIPSLQSSLL-----GQ-----NOKVELACHL--LSTLTT 62
QY 71 EKMPMTSPSCNSCEL-----GIQNLSEKVSQGMALIGSSGGRASLDVITG-- 120
Db 63 YKYPEHGPVLEVDIOKFOHQALHIMIDKAGEFVSFLPSCGKTLIRIINGLE 122
QY 121 -REHGKIKSGQIMWINGOPSSPOLV-----KCAVHRQHNOLLPLTLVRETLAFIAQML 175
Db 123 KPDYGVYTKG-----ADITLLEAKNHC--GIYFQNYALPMLVNEENITFGDL-- 170
QY 176 PTFESQARDRVEDYIAELRLQCADTRYGNMYVSGSGERRYSIGVOLMNPGLI 235
Db 171 KKKMDAQRQSRQVQLLELLELD-----IAKYPNOLSSGQOQVVALARATIPNDLIL 225
QY 236 LDEPTSGLDSTAHNLVKTLSRLAKGNRLVLSIHOPRSDIFLVDLILMTSG----- 289
Db 226 LDEPLSALDAQVRLKIRSIQTOLEPTIWTIDQEARLSISDRVYVMMNGVLEQID 285
QY 290 TP---YLGAQHMYOYFAIGPCPRISNPADFYDLSIDRRSREBELATREKASLA 346
Db 286 TPNHITYKQTOFVAEFIGTMNPLKATCAVNPQLEV--LGFIPMLNLEQOKIKAGENYS-- 342
QY 347 ALELEKVRDLDELKMAEFTKDEDTCEVSSVTPIDTNCPLSPTKMGA-----YQOFTT 401
Db 343 GFRPEAVELVDGSGKRESLTL-----PVAVLSTEEFLGAKRRLLCAIHIDIGIOAKH 394
QY 402 LIRQIISND-----FRDLPTLLIH 420
Db 395 LQIEIENKFKSLQELMFKVNPOLIH 422

RESULT 11

US-09-252-991A-18872
Sequence 18872, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18872
LENGTH: 345
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18872

Query Match 6.4%; Score 225; DB 4; Length 345;
Best Local Similarity 26.2%; Pred. No. 3,7e-16;
Matches 67; Conservative 53; Mismatches 112; Indels 24; Gaps 5;

QY 88 IONLSFKVRSQGMALIGSSGGRASLDVITGRHGKIKSGQIMWINGOPSSP----- 141
Db 31 LQPTRLNIOAGQIFGLIGSGAKSTLILTL---NLEPSSGRLIVBEDVATLDAEGL 87
QY 142 QLVKCAVAVRHQHNOLLPLTLVRETLAFIAQMLPRTFSQARDRKVEDYIAELRLQCA 201
Db 88 RPRRQVGMIFQHFNLISKTVADNIAM--PLRLAGGFRAVDARVSELLARVGLSDHA 145
QY 202 DTRVGNMYVRSIGSGERRRYSIGVOLMNPGLIILDEPTSGLDSTAHNLVYTLISRLAKG 261
Db 146 -----KRYPAQISGGQKQVGIARALACRPSILIDEDASALDPOTVTSVILQDLLEINRE 200
QY 262 NRLVLISLHQRSDIFRLFDVLVLTMTSGPTIYGAQHMYOYFAIGYCCPRYSNPADRY 321
Db 201 LKLTIVLITLHMDVIRRCVQDVAVNDGCAIVDQGVADVFLH-----POHPTTRREV 252

QY 322 VDLTSIDRRSREDELA 337
Db 253 FEAERVEDERHDDFA 268

RESULT 12

US-08-665-259-27
Sequence 27, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klingert, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1457 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-665-259-27

Query Match 6.3%; Score 222.5; DB 3; Length 1457;
Best Local Similarity 22.6%; Pred. No. 9e-15;
Matches 78; Conservative 68; Mismatches 138; Indels 61; Gaps 8;

QY 81 QNSCEGIONLSFKVRSQGMALIGSSGGRASLDVITGRHGKIKSGQIMWINGOPSS 140
Db 22 KDKKTLAKKLKLSLLENQVVSFLGNAGKTTTSLILGL---FPPTGSAITYCHDIR 78
QY 141 POL--VRKCAVAVRHQHNOLLPLTLVRETLAFIAQMLPRTFSQARDRKVEDYIAELRLR 198
Db 79 TEMDEIRKMLGMCQPNVLFDLRLVBEHMLFYSRL--KSMQDELRKTDKMIEDLEL- 134
QY 199 QCADRVGMVYRSGSGERRRYSIGVOLMNPGLIILDEPTSGLDSTAHNLVYTLISRL 258
Db 135 ---SNKRSLVQTLISGGMKRLISVAIAFVGSRAIILDEPTAGVDPYARRAILMDLILKY 190
QY 259 AKGNRLVLSLHQRSDIFRLFDVLVLTMTSGPTIYGAQHMYOYFAIGYCCPRYSNPADRY 310
Db 191 KRGKRTILSTHMDL--LGDRAITLISHGKLKCCGSPFLKGAIXDGYRLTLVKQPAE 248
QY 311 -----CPRYSNPADRYVD-----LTSIDRRSR 332

Db 249 PGTSOEPGLASSPGCPEPLSSCSEPVSQFIRKRVASSLTVSDTSTELSTYILPSEAVK 308
Qy 333 EOEIATREKAQSLAFLFEKVRDLDLFLMKAETKOLDPQVSS 377
Db 309 AFERLFOGLEHSLDALHLSRGLMDTTLVEVFLKVSSEDSLENS 353

RESULT 13
US-08-762-500-27

: Sequence 27, Application US/08762500
: Patent No. 6030806
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burn, Timothy C.
: APPLICANT: Connors, Timothy D.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
: NUMBER OF SEQUENCES: 83
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,500
: FILING DATE: 09-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/665,259
: FILING DATE: 17-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10469
: FILING DATE: 17-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: IGS-9.3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 872-8400
: TELEFAX: (508) 872-5415
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1457 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-762-500-27

Query Match 6.3%; Score 222.5; DB 3; Length 1457;
Best Local Similarity 22.6%; Pred. No. 9e-15;
Matches 78; Conservative 68; Mismatches 138; Indels 61; Gaps 8;

Qy 81 QNSCEGLONLSKRVSGOMLAIIGSSGGRASLDVITGRGKGIRKSGQIWINQOPSS 140
Db 22 KDKKRLAKLKLKSLINLENOVVSFLGHNGAGKTTTMSILNGL--FPPTSGSATIYCHDR 78
Qy 141 POL--VRGCAVHRQNOHLPNTVRETLAFLAQMRLPRTFSQAQDRKVEDVIALRLR 198
Db 79 TENDETRNKILGKCPQNPVFLPDRLLVEHMFYSRL--KSMQOEIRKEDDKIEDLEL- 134
Qy 199 OCADTRVGNMYVNGLSGGERRRVSGVQLMNPGLILDEPTSGIDSPFAHNLVKTLRL 258

Db 135 ----SNKRHSLVOTLISGCMKRLVAIAFVGSRAIILDEPTAGVDPIARRAINDLILKY 190
Qy 259 AKGNRLVILSLHOPSDIFRLDVLMT-----SCTPIYGAQHVVQYFAYGP-- 310
Db 191 KPGRTILSLTHHMEADL--LGDRIAISHGKLKCCGSPFLKGAIXGYRLTLVKQPAE 248
Qy 311 -----CPYSNPADEFYVD-----LTSIDRRSR 332
Db 249 PGTSOEPGLASSPGCPEPLSSCSEPVSQFIRKRVASSLTVSDTSTELSTYILPSEAVK 308
Qy 333 EOEIATREKAQSLAFLFEKVRDLDLFLMKAETKOLDPQVSS 377
Db 309 AFERLFOGLEHSLDALHLSRGLMDTTLVEVFLKVSSEDSLENS 353

RESULT 14
US-08-311-731A-209

: Sequence 209, Application US/08311731A
: Patent No. 6583266
: GENERAL INFORMATION:
: APPLICANT: SMITH, DOUGLAS
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
: TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
: TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
: NUMBER OF SEQUENCES: 411
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: WOLFE, GREENFIELD & SACKS, P.C.
: STREET: 600 ATLANTIC AVENUE
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/311,731A
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: GATES, EDWARD R.
: REGISTRATION NUMBER: 31,616
: REFERENCE/DOCKET NUMBER: C0044/7125
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/720-3500
: TELEFAX: 617/720-2441
: INFORMATION FOR SEQ ID NO: 209:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 396 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Mycobacterium leprae
: US-08-311-731A-209

Query Match 6.3%; Score 219.5; DB 4; Length 396;
Best Local Similarity 28.5%; Pred. No. 2e-15;
Matches 69; Conservative 52; Mismatches 84; Indels 37; Gaps 9;

Qy 88 IONLSKRVSGOMLAIIGSSGGRASLDVITGRGKGIRKSGQIWINQOPSSPOLVR-K 146
Db 24 VHDLSITVADGEFLILIGSGCGKTTLLNMIAGL---EDISSGELRIDGRRVNEKAPKOR 80
Qy 147 CVAHVROHQLDLNLTIVRETLAF---IAQMRILPRTFSQAQDRKVEDVIALRLRQCAD 203
Db 81 DIAMVQSYALVHYHMYVRQNIAPFLAKYK-----KAELAKVSETAQIIDLTLDLDR 134
Qy 204 RVGNMTYVRLGSLGGERRRVSGVQLMNPGLILDEPTSGIDSPFAHNLVKTLRLAKGNR 263

Db 135 KPSQ-----LSGGGRVAMGRAIVRHAFIMDEPLSNLD---AKLRVTRGELIARLOR 186
QY 264 ---LVLSLHQPSPDIFRLEFDVLMTSGTPYIAGAAOHWOYFALIGYPCPRYSNPADF 320
Db 187 RLGAFTVYVTHQOTEAFTGLGDRVYVVRSG-----VVG---QIGTPDELXYERPVNL 233
QY 321 YV 322
Db 234 FV 235

RESULT 15

US-09-252-991A-21751
Sequence 21751, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21751
LENGTH: 676
TYPE: PR
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21751

Query Match

Best Local Similarity 6.3%; Score 219.5; DB 4; Length 676;

Matches 75; Conservative 53; Mismatches 97; Indels 57; Gaps 10;

QY 46 TLEVRDL--NYQVDLASQVPMFEOQLAQFMPTSPSCNSCELGIONLSFKVRSQMLAI 103
Db 293 TELRNHLKSYSGSLADT-----LRDIRLKTIDSGEFLIL 326
QY 104 IGSSCGGRASLDVITGRGHGKIKSGOIWINGOPSSPOLVR-KCVAHYRQHNOLLPLNT 162
Db 327 VGPSCGKSTLMNCIAGL---ESITGAILVDGQDISGMPKRDIAWVFGSYALYPTMS 383
QY 163 VRETLAFTAOMLPRTFSQAQRDRVEDYIAELRLQCADTRVGMAYVAGLSGGERRRYS 222
Db 384 VRENIAFGIKT---RKMPQALIDEVAVARILQIEHLERRPSQ-----LSGGGOORVA 435
QY 223 IGVOLLMPGILILDEPTSGIDSFYAHNLVKTLSRLAKGNRLVLSLHO--PRSDIFRLF 280
Db 436 MGRALARBRPKVYLFDEPLSNLD-----AKLRVEMRTELKLMHQRLKTTVYVTH 484
QY 281 DLVLTMTSGTPYIAGAAOHWOYFALIGYPCPRYSNPADFVY 322
Db 485 DQIEAMTGLGDRVAV-MKDGVIQOF---GTPQOIYNDPANTLFV 522

Search completed: July 25, 2003, 17:16:09
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:08:49 ; Search time 43 seconds

(without alignments)
1505.152 Million cell updates/sec

Title: US-09-989-981a-8

Perfect score: 3506

Sequence: 1 MAGKAABERGLPKGATPQDT.....FNVLYVSLRFIKRPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_76:***
2: PIR2:***
3: PIR3:***
4: PIR4:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	735.5	21.0	725	2 C84423	probable ABC trans
2	723.5	20.6	646	2 C86441	probable ABC trans
3	700	20.0	609	2 E96742	probable ABC trans
4	668.5	19.1	737	2 T46101	ABC transporter-11
5	656	18.7	687	1 JCFEWM	white protein - fr
6	653	18.6	656	2 S77860	brain multidrug re
7	653	18.6	635	2 T08934	probable membrane
8	651.5	18.6	635	2 T08934	hypothetical prote
9	618	17.6	638	2 G02068	white homolog - hu
10	595	17.0	687	2 D96553	hypothetical prote
11	591	16.9	725	2 T47652	ABC transporter-11
12	590	16.8	559	2 B88474	protein C05D10.3
13	589.5	16.8	720	2 T47648	ABC transporter-11
14	586.5	16.7	708	2 T47650	ABC transporter-11
15	579	16.5	658	2 T31958	hypothetical prote
16	575.5	16.4	678	2 H96552	hypothetical prote
17	573.5	16.4	646	2 JCF777	ATP binding cassel
18	562.5	16.0	1049	1 S19421	ATP-dependent penm
19	561.5	16.0	739	2 T45891	ABC transporter-11
20	561.5	16.0	755	2 G84791	probable ABC trans
21	556.5	15.9	740	1 T02567	probable ABC trans
22	550	15.7	649	2 A84509	probable ABC trans
23	548	15.6	662	2 T47649	ABC transporter-11
24	547	15.6	695	2 B96573	protein F12M16.17
25	539	15.4	695	2 T21109	hypothetical prote
26	534	15.2	633	2 T19189	hypothetical prote
27	533.5	15.2	659	2 E86313	hypothetical prote
28	527	15.0	1443	2 T02491	probable ABC trans
29	525.5	15.0	577	2 T04229	ABC-type transport

30	524	14.9	608	2 T34391	hypothetical prote
31	521	14.9	639	2 C88839	protein C106.5 [1
32	511.5	14.6	547	2 T31543	hypothetical prote
33	511	14.6	610	2 T19333	hypothetical prote
34	506.5	14.4	1423	2 A86289	probable ABC trans
35	480	13.7	1450	2 T45888	ABC transporter-11
36	474	13.5	1451	2 B86286	ABC-type transport
37	467.5	13.3	1420	2 T02644	ATP-binding cassel
38	466.5	13.3	1426	2 T30567	probable transport
39	464	13.2	1564	2 S55517	brown protein - fr
40	463.5	13.2	675	2 S55517	probable ABC trans
41	461	13.1	1469	2 H96622	protein putative A
42	452	12.9	1435	2 D96693	probable ABC trans
43	446	12.7	1450	2 A84780	probable ABC trans
44	440.5	12.6	1413	2 G84790	probable ABC trans
45	437	12.5	668	2 S55023	brown protein - fr

ALIGNMENTS

RESULT 1	C84423	probable ABC transporter [imported] - Arabidopsis thaliana
C:Species:	Arabidopsis thaliana (mouse-ear cress)	
C:Date:	02-Feb-2001	#sequence, revision 02-Feb-2001 #text, change 02-Feb-2001
C:Accession:	C84423	
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tai, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ver Natute 402; 761-768, 1999		
A:title:	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number:	A84420; M01D:20083487; PMID:10617197	
A:Accession:	C84423	
A>Status:	preliminary	
A:Molecule type:	DNA	
A:Residues:	1-725 <SNO>	
A:Cross-references:	GB:AE002093; NID:g4262239; PIDN:AAD14532.1; GSPDB:GN00139	
C:Genetics:		
A:Gene:	At2g01320	
A:Map position:	2	
Query Match	21.0%; Score 735.5; DB 2; Length 725;	
Best Local Similarity	30.0%; Pred. No. 3.6e-50;	
Matches	186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;	
QY	75 WTSPSC-----QNSCELGIONLSFKVRSQGMALITGGSGCGRASILDVYTG-----RG 122	
DB	72 WRNITCSLDKSSKSVRFILKNVSGEAKPGRLLAIMPSSGKTYTLINVLAVGLSLSPRL 131	
QY	123 HGGRKSGQIWIINGOPSPOLVVKCAVYRHOHLQPLNLTVRETLAFIAQMRPPTFSQA 182	
DB	132 H-----LSGLELVNGKRPSSKAYK--LAFVRQEDLFFSOLTVRRTLSFAELQPELSSAE 185	
QY	183 QNDKREVDYIAELRQCADTRYGNMYVGLSGGERRRVSIGVQLMNPGLILDEPTSG 242	
DB	186 ERDEYVNNLLKIGLVSCADSCGDAKVGKISGGEKRLSLACELIASVPIADBPPTG 245	
QY	243 LDSEFHNHVKTLRSKAGNRLVLTSLHOPRSDIFPLFDVLMTGCTPIYIG-AAQHW 301	
DB	246 LDFQAEKVMETLQKADGHTVYICIHOPRGSVYAKFDIVLTGTLVYVGPACKEPL 305	
QY	302 QYFTAGYPCPRYSNPADFYVDLTSIDRSRQELATRRKAOSIALFLEKRDLDLFLM 361	
DB	306 TYFGNNGFCPEHPVPAEFLADLISDVSSSEYVSSQKRVAVLADVFQR----- 356	
QY	362 KAEFDLDLDEDTVESSV---PLDINCLPSPRK-----MPGAVOQFTTLIR 405	
DB	357 -----SSSVLATPLS---MKEETKNGMRPRKAIIVERTDMMHOFLLKR 400	
QY	406 QISNDFRDLPTLLINGAEACLSMTIGLFGHSGIQISFMOTALLMFIAGLIPFNVL 465	
DB	401 AMMQASRQGPITKVRARMSVASATVIFGVSFMRGKSQTSIQDRIGLL-OVAI---NTAM 456	

SVV EDICVSS- - - - -VIPEDINCLPSPTKMPGAVQDFITLIRKQI-SNDFRDLPTLLIHGAEA 42

[illegible]

```

Db      240 RDAMAFESVGSFAPFPMNPAFLDIA--NGVCTDGTREKPNVOTLVATY----- 292
QY      355 DLDPLMKAEFTKDLDEPCVSSVPLDTNCLPSPKMGAVQOFTTLI----- 403
Db      293 ---DPLLAPOVK-----TCIEVSHFPD--NAFVATRVNGG--GTTTCAATFSQLCILL 341
QY      404 -----RQISNDRFDLPLLIHGAECALMSFTIFLFGHSGIOLSEMDTALLPMI-- 455
Db      342 HRLKRRHESFD-----LRIPOVVAASITCIGIMW--HSDYR--DVHDIRGLLFFPISI 392
QY      456 --GALIPFNVLIDVSKCYSERAMLYELEDGLTYTGPFPAKIIIGELPEHCAYIIITICM 513
Db      393 FMGVLPSPFNAYFTF---PQERAITRERASGMITLSITFMAHVASLSMELVLPASELT 448
QY      514 PTYMLANIRPGLQPLFLHVLWLVVFCRIMATAAALLPTFHMASFNSALNSFYLA 573
Db      449 FTYMAYVLRPGIVPLLTSLVLLVLAOGIGLALGAINDAKASTIVYTMALAFVLT 508
QY      574 GGFMINLSSMTVPA---WISKVSFLKRCFEGMLKIQPSRRYKMPGLNLTIVASGDKIL 630
Db      509 GGYIVN-----KVPSGMVMKRYVSTFYCYRLVAIOYQ-----SGEILL 548
QY      631 SAMELDSYPLXA-----IYLVIGLSG-----GFWLYVYSLREIK 666
Db      549 RMLGCDSKGKGASATSAAGCRFVEEYIGDVGMTSVGLFLMEFGYRLAYIALRIK 608

```

RESULT 4

```

746101
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T25B15.80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46101
R:Alcaraz, J.P.; Clabault, G.; Cortet, A.; Maché, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223021
A:Accession: T46101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <ALC>
A:Cross-references: EMBL:AL132972
A:Experimental source: Cultivar Columbia; BAC clone T25B15
C:Genetics:
A:Map position: 3
A:Intons: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3
A:Note: T25B15.80

```

```

Query Match      19.18; Score 668.5; DB 2; Length 737;
Best Local Similarity 28.08; Pred. No. 7,7e-45;
Matches 189; Conservative 134; Mismatches 258; Indels 93; Gaps 16;

QY      25 DRLFSSSDNSLKYFTYSGOPN---TLEVRDLNVOYDLASQVPMFEQLAQKRMPTSPSCQ 81
Db      120 DLEDEFAATSSVVKFAEPTFPYIKFDITYKY-----TTRGMT 160
QY      82 NSCELAGION-LSEFVRSGOMALIISSGCRASILDVITGRGHGKIKSGOIMWNGPSS 140
Db      161 SSSEKSLINGISGSAVYGEILLALMGPSGSKTILLNLGSRFNOOI--GGSVSINDPYIS 219
QY      141 POLVRCAVAVRHQNDLPLNLYVRETLAFLAOKRLPTFSQAORKEVEDVIAELRLROC 200
Db      220 KHLKTR-IGFVTDDVDFPLHFKETLTYALRLPRTLREQEKEGSAASVIGLEERC 278
QY      201 ADRVRGMVYRGISGGRRRVSYGVOLAMPGLILDEPSGDSFLAHLVYTLSLAK 260
Db      279 QDMIGGSFVGVSGGRKVCIGNEIMTNPSSLILDEPSSDSTALKTIVOMLHCIAK 338
QY      261 GNLVLISLHQPDSIDFLRLFDVLMTSGPIYGAAGHVOVFTYAGCYCPRYSNADF 320
Db      339 AGRTIYTTIHQPSSRLFHRDKLVLSRGLYFGAASEMASTFSSIGCSPLLAAMNAER 398
QY      321 YVDLTISDRSRQELATYRKAQSL-AALFLERVRLDPLFMKAEKTKLDEDEPCVSSVT 379

```

```

Db      399 LLDLVNGNNDISVPSALKKKIIRLELYANVK-----CDVEQYILE--AKTQIA 450
QY      380 PLDTNCLPSPKMGAVQOFTTLIRQISNDRFDLPLLIHGAECALMSM----- 429
Db      451 VMEKKILAMPVLDDEVRKIMTICPKREMG-----LSMWEQCLSLRQIKERRHY 501
QY      430 -----TIGFLYFGHGSIQLSFMDTALLFMIGALLIPFNVLIDVSKCYSER 475
Db      502 FSWLAVTQVLSAIIIGLMM--QSDTTSQRPFRSGLLEFVAVGWFPEFTALIFPPQER 560
QY      476 AMLYELEDGLTYTGPFPAKIIIGELPEHCAYIIIGMPTYMLANIRPGLQPLFLHVLV 535
Db      561 AMLSKERESNMVRLSAIFYAFRTSDLPDLILPELVVYVFMAGRLRAESFLLSVLT 620
QY      536 WLIVFCRIMATAAALLPTFHMASFNSALNSFYLAGGFMINLSSMTVPA---AMISK 592
Db      621 FLCIVAAOGIGLALGASLMDLKKATTLASVVTWTFMAGGYEVK-----KVPFIAMIRF 675
QY      593 VSEFLMCEGMLKIQPSRRYKMPGLNLTIVASGDKILSAMELDSYPLXAIYVIGLSG 652
Db      676 MSENTHYKLVKVOYE-----ELMESVNGEIESGLK---EVSALVAMIT----- 718
QY      653 GFWLYVYSLREIK 666
Db      719 GYRLVAVFSLRMRK 732

```

RESULT 5

```

FYEFW
white protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1990 #sequence_revision 17-Feb-1995 #text_change 19-Jan-2001
C:Accession: S08635; S07263; S10240
R:Pepling, M.; Mount, S.M.
Nucleic Acids Res. 18, 1633, 1990
A:Title: Sequence of a cDNA from the Drosophila melanogaster white gene.
A:Reference number: S08635; MUID:90221897; PMID:2109311
A:Accession: S08635
A:Molecule type: mRNA
A:Residues: 1-687 <PEP>
A:Cross-references: EMBL:X51749; NID:98825; PIDN:CAA36038.1; PID:98826
R:O'Hare, K.; Murphy, C.; Lewis, R.; Rubin, G.M.
J. Mol. Biol. 180, 437-455, 1984
A:Title: DNA sequence of the white locus of Drosophila melanogaster.
A:Reference number: S07263; MUID:85134865; PMID:6084717
A:Accession: S07263
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCVTAD', 30-334, 'ITLHLSYPAWVPSVLPTRIRRTYRCMPLCFPGSSPI
A:Cross-references: EMBL:X02974
A:Experimental source: Strain Canton S
R:O'Hare, K.
submitted to the EMBL Data Library, June 1985
A:Reference number: S10240
A:Accession: S10240
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCVTAD', 30-687 <ONH2>
A:Cross-references: EMBL:X02974; NID:910873; PIDN:CAA26716.1; PID:910874
A:Experimental source: Strain Canton S
C:Genetics:
A:Gene: white; w
A:Cross-references: FlyBase:FBgn0003996
A:Intons: 24/3; 116/1; 334/2; 439/3; 483/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
F:113-137/Domain: ATP-binding cassette homology <ABC>
F:130-137/Region: nucleotide-binding motif A (P-loop)
F:261-265/Region: nucleotide-binding motif B
F:67,93,472,554,651/Binding site: carboxydrate (Asn) (covalent) #status predicted

```

```

Query Match      18.78; Score 656; DB 1; Length 687;
Best Local Similarity 30.38; Pred. No. 6,9e-44;
Matches 178; Conservative 113; Mismatches 265; Indels 32; Gaps 10;

```

OY 88 IONLSPFRVSGOMLAIIGSSGGRASLLDVTGKHG--KIKSGQIWMGSPPOLVR 145
 Db 113 LKAVCGVAYPGELLAWSSGACKTLLNLALFASPGIOVSPSGMLNQGPDAREMQ 172
 OY 146 KCAVAVRQHNOLLPNLVRETLAFIAOMRPFPSAORDKRYDVIAELRLGCATRV 205
 Db 173 ARCAVVOODDLFGSLFARHLLFOAMVRPRHLLYRORARVQVIOELSLSCQHTI 232
 OY 206 G-NMYVRLSGGERRRVSIGVOLLMPGILLDEPTSGDSTFANLTKLSRLAKNRL 264
 Db 233 GVGGRVYKLSGGERRKRLAFASDALTPDLICEPTSGDSTFANVVOYLKRLSKQK 292
 OY 265 VILSHOPRSDIRLDVILMTSGPTIYGAAGHVOYFTALGTPCPKRSNPADFTVDL 324
 Db 293 VILTHOPSELEFELDKILLAMEGRVAFGTPESEADVFSSYGAQCPMYNPADFTV 352
 OY 325 TSIDRRSRQELATREAGSLALF-LEKY-RDLDPFLMAEKTKDDEDCESSVPLD 382
 Db 353 LNV---VPGREISRRRIAKICNFAISKVARDMEQL---ATKMLEK-----PLE 397
 OY 383 TNCLESP---TKMPAVOQFTLLIRROISNDFRDLPTLLIHGAECALMSMTIGFLYFG 438
 Db 398 ---OPENGYTKATWQFRAVLMRSMLSVLKEPLVAVRLIQTMAILLGLIFLGQ 452
 OY 439 GSTQLSFMDPAALLPMIGALIPPNVILDVISKYSEBAMLYELEDEGLYTGPFPAKIL 498
 Db 453 QLTQVGMNINGAIFLEFLNMTFQNVFATINFTSELPEVPMREARSRLNCDYFLCKTI 512
 OY 499 GELPEHCAYIIIGMPTMYLANLRPGLOPFLHFLVWLVVPCCRIMALAALLPFFHM 558
 Db 513 AELPELTVPLVYTAALTAIPMIGALHFNCLALVTLAVANSTSEGYILISCASSSTM 572
 OY 559 ASEFSSNALYNSFYLAGFMILNLSLMTVPAMISKVSFLMCFEGIMKIQPS---RRTYKM 615
 Db 573 ALSVGPVLIIPFLFGFGLNSGSPVYLKMLSYLSMFRANEGLLINOMADVEPEISC 632
 OY 616 PLGNLTIVASGDKILSAMELDSTPLAIVLIVIGLSGGGVNLYYSLR 663
 Db 633 TSNNTCPSSGKVILETFLNFSADPLDLYGVALILVSEFVLAFLMR 680

RESULT 6

JC7860
 brain multidrug resistance protein, BMDP - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #text_change 31-Mar-2003
 C:Accession: JC7860
 R:Eisenblatter, T.; Gallia, H.J.
 Blochem. Biophys. Res. Commun. 293, 1273-1278, 2002
 A:Title: A new multidrug resistance protein at the blood-brain barrier.
 A:Reference number: JC7860; M0ID:22050127; PMID:12054514
 A:Accession: JC7860
 A:Molecule type: mRNA
 A:Residues: 1-656 <EIS>
 A:Cross-references: GB:AJ420927
 A:Experimental source: brain
 C:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) super
 exclusion of xenobiotics from the brain and participates in drug transport across the bl
 C:Genetics:
 A:Gene: bmdp

Query Match 18.6%; Score 653; DB 2; Length 656;
 Best Local Similarity 27.6%; Pred. No. 1.1e-43;
 Matches 192; Conservative 133; Mismatches 270; Indels 100; Gaps 20;
 OY 18 OPTSGLODRLFSESNSLYTYSQPNLEVRDLNVOYDASQVWFQOLAQFKMPWIS 77
 Db 15 RNTNGL-----PGSSNELKTSAGG--VLSFHDICRYKRVKSGFL----- 54
 OY 78 PSCQNSCELGI-QNLSFVRSQOMLAIIGSSGGRASLLDVTGGR--GHGKIKSGQIWI 134
 Db 55 -CRATVEKEILTININGIMKPG-LNALIGPTGGKSSLLDVAARDPHG-----LSQDVI 108

OY 135 NGOPSSPOLVKKC-VAHVROHNOILLPNLIVRETLAFIAOMRPFPSAORDKRVEDVIA 193
 Db 109 NGAPRANF--KCMGSYVQDDVVMGLTYRENIQFSAALRLPTTMTNHEKNEINNAVIO 166
 OY 194 ELRLRCQADRVRGMMYVRLSGGERRRVSIGVOLLMPGILLDEPTSGDSTFANLTVK 253
 Db 167 ELGLDKVAADSKVGTQFIRVSGGERRKRTSIAMELITDPSILFIDEPTTGDSSANVLL 226
 OY 254 TLSRLAKGNLVLVISHOPRSDIRLDVILMTSGPTIYGAAGHVOYFTALGTCPR 313
 Db 227 LKRMKSQGRITIFSIHQPRYSIFKLPDSTLLASGMLRHGPARLEGFASIGYCEP 286
 OY 314 YSNPADPYVDLTS-----IDRRSRQELATREK-----AQSLA-----LFLEK 352
 Db 287 YNPADPEFLVINDSSAVVLSRADRGQDEPEPEPEKPTPLDKLAAYTNSFFKDT 346
 OY 353 VRDLDFLMAETKDEDCESSVTPDLTNCILPSPMKGAVOQFTLLIRROISNDFR 412
 Db 347 KVELDQFSGGRKKR---KSSVYKEVYTTSSFC-----HQLRMISRSRGNLLG 391
 OY 413 DLPTLLIHGAECALMSMTIGFLYFGHSIQLSFMDPAALLPMIGALIPPNVILDVISKY 472
 Db 392 NPOASVQIITVITILGLVIGALIFYDLKNDPSGIONRAGVLEFL-----TTNOCF 440
 OY 473 S-----ERAMLYELEDEGLYTGPFPAKILGE-LPEHCAYIIIGMPTMYLANL 521
 Db 441 SSVASVABLVYERKFLFHEIYSGYRRSSIFFGKLSDLLPMRMLSIITCTIYFLGL 500
 OY 522 RPLGLPFLHFLVWLVVPCCRIMALAALLPFFHMASFSSNALYNSFYLAGFMILNLS 581
 Db 501 KPAVGSFFIMFTLMVAVSASSMALAIAGQSVSVATLMTISFPMHIFGLLVNKK 560
 OY 582 SLMTVPAMISKVSFLMCFEGIMKIQPSRRTYMPIGNLT-----IAVSGDKIL--SA 632
 Db 561 TVVPMLSMLQYFISIPRGFSALQYNELGNF--CPGLNVTNTNTCSFALCTGAEYLENOG 619
 OY 633 MELDSTPLAIVLIVIGLSGGFMVLYYSLRFIKQ 667
 Db 620 ISLSANGLMQNHVALACMAYIFLTIVAKILLILKK 654

RESULT 7

S77690
 Probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothet
 C:Species: Saccharomyces cerevisiae
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
 C:Accession: S77690; S66767; S66768
 R:Alexandraki, D.; Katsoulou, C.; Tzeremia, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66756
 A:Accession: S77690
 A:Molecule type: DNA
 A:Residues: 1-1294 <ALE>
 A:Cross-references: EMBL:Z74816; MIPS:YOL075c
 A:Note: this is a revision to the sequence from reference S66756
 A:Accession: S66767
 A:Molecule type: DNA
 A:Residues: 1-179, 'TRTGVFLVYKRED' <ALM>
 A:Cross-references: EMBL:Z74816
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Note: this was assumed to be protein YOL074c
 A:Accession: S66768
 A:Molecule type: DNA
 A:Residues: 200-1294 <ALF>
 A:Cross-references: EMBL:Z74817
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Note: this was assumed to be the complete sequence of protein YOL075c
 A:Cross-references: SGD:S0005435

A:Map position: 15L
 A:Note: Y01075C
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:376-392/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:606-622/Domain: transmembrane #status predicted <TM4>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1142-1056/Domain: transmembrane #status predicted <TM5>
 F:1125-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.6%; Score 653; DB 2; Length 1294;
 Best Local Similarity 30.1%; Pred. No. 2.9e-43;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;

QY 88 IONLSPKVRSGOMLAIIGSSGGRASLDVITGRGKGKIKSGQI-----132
 Db 45 VMTFSMDLPESGVAAVWAGSGSKTTLNVLASKISGILTHNGSIRVLEDTGSEPNETE 104
 QY 133 ---WINGOPSSPOLVRKCAVHRQHNLPLNLTVEETLAFIAQMRLPRTFSQAORDKR- 187
 Db 105 PKRAHLDGQ-DHPIDKHVIMAVLPQDDVLSPLRTCHETLKFADKL-----NSSERTKLL 159
 QY 188 -VEDVLAELRLKOCADTRNGNMYRGLSGGERRRVSIGVOLLNPGILLDEPTSGLDSE 246
 Db 160 MVBOLLEELGLKCCADTLVGDNSHRGLSGGERRRLSIGTQMISNPSIMFLDEPTGDAV 219
 QY 247 TAINLVKTLRLAK-GNRVLISLHOPRSDIFRLFDVILMTSGTPIYLGAAGHMYQYF 305
 Db 220 SAEVLVITKLKLKKEGGRFTMSIHQPRSDILFLDQVCLLSKGNVYTCDKMNTPIYFE 279
 QY 306 AIGPCPRYSPADFYVDLTSIDRSREDELATREKASLAFLFEKVRDLDFEIKAEF 365
 Db 280 SIGYHVPOLVNPADYFIDLSVDSRSDKEKAAQSLNSL-----IDHMD-----YERTH 330
 QY 366 KDLDEPTCVESSTPLDPTNCLSPYKMGAVQDFTLLIRQISNDRDLPTLLIHGAENK 425
 Db 331 LQIQAESYI-SNAETIOIONM--TTRLP-FMKQVTLTRNFKLNSDVYTLISTFAEPL 386
 QY 426 LMSMTIGFLYFGHSGISQLSFMOTALLFMIGALIP--ENVILDVISKYSERAMLYELE 483
 Db 387 IIGTCGMIYKPRDKSSISGLRTTACLYASTILOCYLILFDTYRICEODIALYDERA 446
 QY 484 DGLYTTGPYFEFA-KIIGELPERHCAYIIIGMPTMYLANLRPGLOPLHLHVLVAVPCC 542
 Db 447 EGSVPLAFIVARKISLFLSDPAMTIVISITIFHFGLEADARKFFYPFAVEFLQSLSC 506
 QY 543 RIMATAAALLPETHMASFFSNALYNSFYLAGFMINISLMTVPAMISKVSFLRMCFGS 602
 Db 507 SGLSMISVAVSRDFSKASLVGNMTFTVLSMCGCFEYNAKVPYVWIKYIAFTWVSFGT 566
 QY 603 LMKIOPSRR-----TYKMLGNITIVSG 626
 Db 567 LMSSTFTNSYCTTNDLDECLAGNOILEVYG 595

RESULT 8
 T08934
 hypothetical protein F27G19.20 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000
 C:Accession: T08934
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, A.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16519
 A:Accession: T08934
 A:Molecule type: DNA

A:Residues: 1-635 <BNV>
 A:Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20
 A:Experimental source: cultivar Columbia; BAC clone F27G19
 C:Genetics:
 A:Gene: ATSP:F27G19.20
 A:Map position: 4
 A:Introns: 38/3; 253/L; 304/L; 414/3
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 18.6%; Score 651.5; DB 2; Length 635;
 Best Local Similarity 31.1%; Pred. No. 1.4e-43;
 Matches 191; Conservative 104; Mismatches 240; Indels 79; Gaps 19;

QY 30 SESNSLXYFY---SGQPNLEVDLNYQVDLASQVWEFQLOAKFKMPTSPSCNSCEL 86
 Db 17 TNDRSILPFSIFKANNPVLKFEENLVYTVKLDKSGCF-----GKNDKTEERT 65
 QY 87 GIONLSPKVRSGOMLAIIGSSGGRASLDVITGRGKGKIK--SGQIWNIGOPSSPOLVR 145
 Db 66 ILKGLTIVAPGELIAMLFGSGSKTSLTALGVRGEGKGLTGNISYNNKPLS-KAVK 124
 QY 146 KCAVHRQHNLPLNLTVEETLAFIAQMRLPRTFSQAORDKRVEDVIAELRLKOCADTRV 205
 Db 125 RTTGEVYQDDALVPLNLTVEETLFTALRLPNSKKEKIKQAKAVTELDGRCDTII 184
 QY 206 GNNYVRLSGGERRRVSIGVOLLNPGILLDEPTSGLDSEFTAHNLVYKTLRLAKGNRV 265
 Db 185 GGPEFLKGVSGGERRRVSIGVOLLNPGILLDEPTSGLDSEFTAHNLVYKTLRLAKGNRV 244
 QY 266 LISLHOPRSDIFRLFDVILMTSGTPIYLGAAGHMYQYFAIGY-PCPRSNPADRYVDL 324
 Db 245 VTIIRHDP-----SKGNPYFELGSNADYFASVGSPLVERINSDFLDDI 290
 QY 325 TSIDR-----RSREDELATREKASLAFLFEKVRDLDFEIKAEKDKLDEDTG---V 374
 Db 291 ANCKPLLVISCWPSVDSDEQREPKAAVLAVYKTNLDSVINEKQGD---DLCKNR 347
 QY 375 ESSVTPLDPTNCLPS-PTKMGAVQDFTLLIRQIS-----NDRDLPTLLIHGAENK 429
 Db 348 ESS--RYANVTYDMPPTW---WQOCVLLKRLKORRHSFGMKV-----AQIFVSP 397
 QY 430 TIGFLYFGHSGISQLS-FMDPALLFMIGALIPENVILDVISKYSERAMLYELEGGIT 488
 Db 398 LCGLLMW---QYTRISRLQDDIGLFTISSWAFPLPQOFTTPPQERAMLOKERRSSGM 454
 QY 489 TGPYFEAKTIGELPERHCAYIIIGMPTMYLANLRPGLOPLHLHVLVAVPCCRIMATA 548
 Db 455 LSPYFLSRVYGDLPMLILPCTCLVITTYMMAGLNHLANFVYLLVHLVAVVSGGLGLA 514
 QY 549 AAALPETHMASFFSNALYNSFYLAGFMINISLMTVPAMISKVSFLRMCFGSLMKIOF 608
 Db 515 LQALVMDOKSATYLVIMLFTLLAGGYVQ-----HVPFISIMKY-----VSI 559
 QY 609 SRRTY-MPLGNIT 621
 Db 560 GYTTIKLILDGQYT 573

RESULT 9
 G02068
 white homolog - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001
 C:Accession: G02068
 R:Croop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Arcin
 submitted to the EMBL Data Library, August 1995
 A:Reference number: H00769
 A:Accession: G02068
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <CRO>
 A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AACS1098.1; PID:g1314277
 C:Genetics:

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-725 <MON>
 A:Cross-references: EMBL:AL132954
 A:Experimental source: cultivar Columbia; BAC clone T26112
 C:Genetics:
 A:Map position: 3
 A:Note: T26112.10
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12U6.1; ATP-B

Query Match 16.9% Score 591; DB 2; Length 725;
 Best Local Similarity 25.5%; Pred. No. 1.1e-38;
 Matches 175; Conservative 135; Mismatches 273; Indels 104; Gaps 18;

```

QY 44 PRTLEVRDLNVOYDLASQVPEFQDLAQFKMPWTSPPSCONCELAGIONLSFVRSQGLAI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 PYVLFNNLQYDYLTLRRRFGF-----SKONGVKYLLDVGSGASDGLAV 115

QY 104 IGSAGGRASLLDVTYGRGHGKIKSGQIWMINGOP-SSPOLVRKCAVHROHQLPMLT 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 IGSAGGRASLLDVTYGRGHGKIKSGQIWMINGOP-SSPOLVRKCAVHROHQLPMLT 174

QY 163 VRELIATIAOMRLPRTSOAQRKRVDAELRLROCADTRGNMTVRGLSGGERRVS 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 VKETLMFASERFLPRSLSKRKMERVRLIDOLRLNAAANTVIGDEGHRGSGGERRVS 234

QY 223 IGVOILMNPGLILDEPTSGLSDFTAHNLVKTSLRKLAKGNRLVLSLHOPRSDIRFLDL 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235 IGDIIDHPIVLFDEPTSGLSDFTAHNLVKTSLRKLAKGNRLVLSLHOPRSDIRFLDL 294

QY 283 VILMTSGTPYILGAQHMVOYFAIGVPCPRYSNPADFYDLTSIDRSREOELATREKA 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 295 ILLSRKSVFNCSFASLPFGFSDFGRIPEKENISEFALDV-----RELE-GSNEG 347

QY 343 OSIALLFLEKVRDLDFLKK-----AEKRLDED-----TCVSSVTP---LDT 383
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 KALVD-FNEK-----WQONKISLQSAPOFNKLDODRSLSLKEAINASVSRGKLYSG 398

QY 384 NCLSPPTKM-----PGAVOFTTLIRROISNDFDLPTLLIHGAELAMNTIGLY 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 399 SSNSNPTSMETVSSYANPLSETFE-ILAKRYMKMNR-MEVL--GTRIVAVVQGLLA 454

QY 436 FCGISQISLSPMDTALLFMIGALIP--FNVILDVISKYSERAMLYELEDGLYTTGPY 493
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 TVYMKLDHPRGAQERLTLEAFVYPTMFVCCLDNVPEIOERITFLRETHNMYKRSISY 514

QY 494 FATHIIGELPRHCAYIIITGMPYTWLANLRPGLOPFLHLFLVAVVFCRIMALAAAIL 553
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 515 ISHISLVSLLPOLAPSLVFSALTFMTVGLSGLEGFEFYCLLIYASFWSGSSVYTFISGV 574

QY 554 PEFHMASFFSNALYNSFYLAGFMINLSL---WTPAMISKVSFLRQCEGLMKQFS- 609
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 575 PNMICLYMISITYLAVCLLSGTYVNRDRIPFTWT---WPHYSILIKYPEAVLILNEFDD 631

QY 610 -----RRTYKMPGLNLTIAVSGDKIISAMELDS 637
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 632 PSRCFVRGYOVFDSTLLGGVSDSGKVKLETLSTKSLRTKITESTCLRTGSDLLAQGQITQ 691

QY 638 YPLIAYILVIGISGCFMVLVYSLEF 664
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 692 LSKMCDLMTFASGLFEFRILFEYFALLF 718

```

RESULT 12

B88474
 protein C05D10.3 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
 C:Accession: B88474
 R:Anonymous
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
 A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Project
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 19
 A:Accession: B88474
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-559 <STO>
 A:Cross-references: GB:chr.III; PIDN:AAA20989.1; PID:9532111; GSPDB:GN00021; CESP:
 A:Note: similar to D. melanogaster white protein
 C:Genetics:
 A:Gene: C05D10.3
 A:Map position: 3
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 16.8% Score 590; DB 2; Length 559;
 Best Local Similarity 29.3%; Pred. No. 9.1e-39;
 Matches 159; Conservative 98; Mismatches 231; Indels 54; Gaps 11;

```

QY 88 IONLSFVRSQGLAIIGSSGGRASLLDVTYGRGHGKIKSGQIWMINGOPSSPOLVRK 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 LHMVSGABSGKLALILGSSGAGKTYLMLNLTNDVQSGILIDGRANKMKITREM 69

QY 148 VAHVROHQLPMLTVRETLAFIOMRL-PRFESQAQRKRVDAELRLROCADTRVG 206
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 SAFVOQHDMEVGTAREHLOFARLMDGQYSDHEROLRVEQYLTQMLKCADVIG 129

QY 207 -NMVRLSGGERRRVSIGVOLLNPGILILDEPTSGLSDFTAHNLVKTSLRKLAKGNRLV 265
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 130 IPNQLKLSGCKKRLSFASEIILCPKILFCDEPTSGLDAPFMAGHYVALRSLADNMTV 189

QY 266 LISLHOPRSDIRFLDLVILMTSGTPYILGAQHMVOYFAIGVPCPRYSNPADFYDLT 325
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 190 IITIHOPSSHVYSLFNWCLMACGRVYILPGQDAPLFEKCGPCPAYVNPADHLIRTL 249

QY 326 SIDRSREDELATREKASLALFLEKVRDLDFLMAKFEKRLDEDGDC-----VES 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 AVIDSDRATSMKT-----ISKIR--QGLF-----STDLAGSVLAIGNANKLRAS 292

QY 377 SVTPDLTNCPLSPPTKM-----PGAVOFTTLIRROISNDFDLPTLLIHGAELAM 429
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 293 FVTGSDPS---EKTKFEFNODYNASFEQTLAFWRSMLVIRDPNLSVRLQIITAF 349

QY 430 TIGELFEGHSQISLSPMDTALLFMIGALIPENVILDY-----ISKYSRAMLTYE 481
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 ITGIVFF-----QTPVTPATITISINGIN-FNHIRNMNEMLOFPNVPVITATLPVLR 401

QY 482 LEDGLYTTGYPFAKILIGELPRHCAYIIITGMPYTWLANLRPGLOPFLHLFLVAVVC 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 NANGVYRTASVFLAKNIAELPQYIILPIILNTIVYNSGLPFNMYCRASLVITLITNV 461

QY 542 CRIMALAAALPTEFHMASFFSNALYNSFYLAGFMINLSLWTPAMISKVSFLRQCE 601
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 462 AISISYAIVATIFANTDVAMTILPIFYVPIMAFGGFTFDPAISYFKMLSSLYFRYGE 521

QY 602 GL 603
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 522 AL 523

```

RESULT 13

T47648
 ABC transporter-like protein - Arabidopsis thaliana
 N:Alternate names: protein T15C9.80
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C:Accession: T47648
 R:Meves, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z24470
 A:Accession: T47648
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <MEM>
 A:Cross-references: EMBL:AL132970

A: Experimental source: cultivar Columbia; BAC clone T15C9
 C: Genes: 3
 A: Map position: 3
 A: Note: T15C9.80
 C: Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 16.8%; Score 589.5; DB 2; Length 720;
 Best Local Similarity 25.0%; Pred. No. 1.4e-38;
 Matches 173; Conservative 122; Mismatches 289; Indels 109; Gaps 14;

```

QY 44 PNTLEVDLNYOVDLASQVWPEQLAQFKPMWTPSPCONSCELGIONLSFKVSGOMLAI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 54 PFLTSFNNLTNYNVRKRDLDH-----LVPMKRTSESKTKL-LNISGETKDEGLIAV 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 104 IGSAGGRASLDLVITGRHGKIKSGQIWINQSPSPOLYRKCAVHROHOLLNLTLY 163
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 108 LGAAGSGKSTLIDALANRIAKGSLK-GTYLNGEALQSRILAKYISAVYMODLLFPMIT 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 164 RETLIAQMRLEPRTSQOQRKVEDVIAELRLROCADTRVGNMYVGLSGERRRVS 223
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 167 EETLMFAEFRLPRSLPKSKKKRLVOALIDQLIRNAAKTIIGDEGHRGSGGERRRVS 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 224 GVOLLNPGILIDEPSTGLDSTFANLVKTLRLAKGNRLVLSLHOPRSDIFRLFDLV 283
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 227 GIDITHDPYLFIDEPTSGLDSTSAFVYKVKRIKIAESGSIIMSHQPSHRYLSLDRL 286
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 284 LMTSGTPYLYGAQHMOYFTAIQYPCPRYSNPADFYDLTS-----IDRS 331
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 287 IFLSRGHVTEGSPASLPSPFAGFNPDPENENQTEFALDLRELGSAGTGLVEFNK 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 332 RDEELATREKASLAA-----LLEKYRDLDDFLMKAEETKDLEDTCVE-----SSVT 379
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 347 KQOEMKQSNPOTLPPASPMPNLTLK-----EASISISRGKLVSGGSSSVY 396
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 380 PLDTNCPSPKMGAVOQFTLLIRROISNDFRDLPTLLIHGAELMSMTIGFLX----- 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 397 NNGGGLAVPAPANPWIETKILTRKSLNSRQ--PELL--GKRLATVITV-GEILATVF 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 436 -----FGHSGIOLSEMDPALLFMIGALIPFNVLIDVYSKYSERAMLYEEL 482
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 WRLDNSPKGVQERLGFPAFAMSTM-----FYTCADALPVFLQERYIMREI 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 483 EGGLYTGYFFRAKILGELPEHCAYIIIGMPTYLANLRPGLOPFLHFLVLMVYVFC 542
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 499 AYNARRSSYVLSHAIVTPPSLIFLSLAFVTFMVAVLEGGLMGLFYCLIIILASFWSG 558
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 543 RIMALLAAALLPFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 602
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 559 SSFVTFILSGVPHVMGLITVIAIYAILFLFSGFTINRRIPOYWIWPHYLSLVKPYEA 618
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 603 LMKIOFSRRTY-----KMLGLNLTIAV-----SGDKI 629
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 619 VLNQFSPDTECFVRCVQLFDNSPIGLDELTYGMLRLLDVSYSIGMRISSTCLTTCADY 678
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 630 LSMELDSYPLAIYIYVIGLSGGFVAVYVSL 662
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 679 LKQGGVTLQSKMNCILITVGFGLFRILFYLCI 711
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14

T47650
 ABC transporter-like protein - Arabidopsis thaliana
 N: Alternate names: protein T15C9.110
 C: Species: Arabidopsis thaliana (mouse-ear cress)
 C: Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C: Accession: T47650
 R: Kewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A: Reference number: Z24470
 A: Accession: T47650
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-708 <MEM>

A: Cross-references: EMBL:AL132970
 A: Experimental source: cultivar Columbia; BAC clone T15C9
 C: Genes: 3
 A: Map position: 3
 A: Note: T15C9.110
 C: Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

Query Match 16.7%; Score 586.5; DB 2; Length 708;
 Best Local Similarity 26.8%; Pred. No. 2.4e-38;
 Matches 187; Conservative 124; Mismatches 254; Indels 133; Gaps 20;

```

QY 44 PNTLEVDLNYOVDLASQVWPEQLAQFKPMWTPSPCONSCELGIONLSFKVSGOMLAI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 PFLTSFNNLTNYNVRKRDLDH-----LVPMKRTSESKTKL-LNISGETKDEGLIAV 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 104 IGSAGGRASLDLVITGRHGKIKSGQIWINQSPSPOLYRKCAVHROHOLLNLTLY 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 108 LGAAGSGKSTLIDALANRIAKGSLK-GTYLNGEALQSRILAKYISAVYMODLLFPMIT 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 VRETLIAQMRLEPRTSQOQRKVEDVIAELRLROCADTRVGNMYVGLSGERRRVS 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 167 EETLMFAEFRLPRSLPKSKKKRLVOALIDQLIRNAAKTIIGDEGHRGSGGERRRVS 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 223 IGVLINPGILIDEPSTGLDSTFANLVKTLRLAKGNRLVLSLHOPRSDIFRLFDLV 282
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 227 GIDITHDPYLFIDEPTSGLDSTNAFVYKVKRIKIAESGSIIMSHQPSHRYLSLDRL 286
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 283 VLMSTGTPYLYGAQHMOYFTAIQYPCPRYSNPADFYDLTSIDRSRDEELATREKA 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 287 IFLSRGHVTEGSPASLPSPFAGFNPDPENENQTEFALDLRELGSAGTGLVEFNK 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 343 OSLAALFLKVRDLDDF--LW-----KAETK--DLDDPTCV-----ESSVTP 381
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 337 -----EGTDVLFENKQWONOTARATTSRVSLEKALIASYRGKLVSGGSGANPT 388
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 382 DTNCPSPKMGAVOQFTLLIRROISNDFRDLPTLLIHGAELMSMTIGFLX----- 431
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 389 SMETVSYANMP--LAETFLAKRYIKNIRPELIGMIGVMTVG-----LLAVYMR 442
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 432 -----GELYFGHSGIOLSEMDPALLFMIGALIPFNVLIDVYSKYSERAMLYEEL 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 443 LINTPRGAOERMGFFAFAMSTM-----FYCCADNIPVFLQERYIMREI 483
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 YELEDGLYTTGYFFRAKILGELPEHCAYIIIGMPTYLANLRPGLOPFLHFLVLMVY 539
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 484 RETTHNAVYTSYVLSHAIVTPPSLIFLSLAFVTFMVAVLEGGLMGLFYCLIIILASFWSG 543
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 540 FCCRIMALLAAALLPFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 599
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 544 WSGSSIVTFISGLIPVMSYMTIAYLSYCLLGGFYINRRIPLXYWIWPHYISLVKXP 603
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 600 FEGIMKIQP---SR---RYKMPGLNLTIAV-----GKILSAMELDST 638
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 604 YEAVLINEFDPSRCVKGQVDFDTLAEVSHVMKVKLLDPLSGSLGKRTISTCLRTG 663
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 639 P-----LYATYIYVIGLSG--FVAVLYVSLRF 664
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 664 PDLMQGGITQLSKMDCMLTILAMGLFRILFYLSLIF 701
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15

T31958
 hypothetical protein F02E11.1 - Caenorhabditis elegans
 C: Species: Caenorhabditis elegans
 C: Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
 C: Accession: T31958
 R: Faveillo, A.; Scheet, P.
 submitted to the EMBL Data Library, July 1997
 A: Description: The sequence of C. elegans cosmid F02E11.
 A: Reference number: Z21104
 A: Accession: T31958
 A: Status: preliminary
 A: Molecule type: DNA
 A: Molecule type: translated from GB/EMBL/DBJ

A:Residues: 1-658 <FAV>

A:Cross-references: EMBL:AF016661; PIDN:AAB66050.1; GSPDB:GN00020; CESP:F02E11.1

A:Experimental source: strain Bristol N2; clone F02E11

C:Genetics:

A:Gene: CESP:F02E11.1

A:Map position: 2

A:Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match

16.5%; Score 579; DB 2; Length 658;

Best Local Similarity 28.3%; Pred. No. 8.5e-38;

Matches 180; Conservative 113; Mismatches 257; Indels 86; Gaps 22;

```
QY 78 PSCONCELGIONLSFKV---RSGQMLAIGSSGGRASLDVYTGKRGKIKSGQIM 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 PECLAVCALPTSSYQISVSGVAEPGEVLALMGSGAGKTLMLNLHLDTNGVEYLGDT 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 INGPSSPOLYRKCAVAVRHQNLPMITVRETLAFLAQMRLPFTFSQARDKNEVDYA 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 VNGKKIKOKROMCAVYQVDLFCGLTYREOLTYAHMRMKNATVQ-QCKEREENVLR 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 194 ELRLQCADTRVG-NMYVRLGSGERRRVSIGVOLNMPGILIDEPSTGDSPTAHNLV 252
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 DMNLTDCONTLIGIPNMAKGISIEKKRIACELTDPKILFCDEPTSGIDAFMASEV 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 253 KTLRLAKGNELVLSHQRSDIFRLDLVLTMTSGTPYILGAQHWOYFTAIQ--YP 310
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 RALIDLANKGKTIIVLHQPSSYFERFMHVCFMATKTYHGAVDRLCPFDKLPDFR 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 CPRSNPADYVLTSLDRSRREDELATR-----EKAQSLALFLEKVRD-LDDEL 362
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 VPESYNPADYVMSISISPEPT-EQEDVTRELYLHEYQNSDYGQMLKRTTAVDERGY 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 AETKDEDCTCVSSVTPDPTNCLPPTKMGANVOQTTLIRQISNDFRDLPTLLHGA 422
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 GDEDDGESRYNSTFGT-----OPEILKRSLRTTFRDPLLRVFA 401
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 423 EACIMSMTIGFLYFGHSIQI---SFMDTAALLFMIGALIPFNVLIDVISKCYSERAM 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 QILATAILVGIV---NMRYVELKGPIONLEGVAMNCARDMTFLFPFSVANVTISELP 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 YELEDGLYTGYPYFAKILGELPEHCAYIIITGMPTWLANLRGLOPFLH---FLV 535
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 REHKSNIYSYEAFLAKSLAELPQYTLPMYGTIIYMAGLVASYSFLVFEVCITLT 518
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 536 MLYVFCGRIMA--LAAALPTFHMASEFNSALYNSFYLAGFMINLSLMTVPAMISKV 593
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 519 WYAVSTAYVAGACIEGDEGLVTF-MPEFVLPML-----VFGFVYVANS--IPVY 569
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 594 SPLRMC---FEGIMAKIQFSRRTYKM-----PLGNLT---AVSGDKILSAMELD 639
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 SFVSWFKHGFEALANQM--KEIDKISGDLINPLNMTTGYCPASDQGLTTRGIDT-P 627
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 640 LYAIYLI-----VIGLSGFVWLYYVSLRFK 666
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 628 LYANVLILFMSFFVYRIIGL-----VALKIRVRFK 658
```

Search completed: July 25, 2003, 17:15:32
Job time : 46 secs

Best Local Similarity 81.8%; Pred. No. 3,5e-216;
Matches 551; Conservative 52; Mismatches 69; Indels 2; Gaps 2;

QY 1 MAGKAAEERGLPGKAPPDITS-GLQDLRFSSSDNSLYFTYSGQPNLLEVRDLYOYDLA 59
DB 1 MAEKTEETQLMNGTVLQDASGLQDLRFSSSDNSLYFTYSGQSNLLEVRDLYOYDLA 60
QY 60 SOVPWFQDLAOFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 119
DB 61 SOVPWFQDLAOFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
QY 120 GRHGGRKSGQIMNGOPSSPOLVRKCVAVRHQNDLPLNLYRETLAIAQRLPRTF 179
DB 121 GRHGGRKSGQIMNGOPSSPOLVRKCVAVRHQNDLPLNLYRETLAIAQRLPRTF 180
QY 180 SOAQRDRVEDVIAELRLQCADTRVGNMRYRGLSGGERRRVSIGVOLLNPGILLIDEP 239
DB 181 SOAQRDRVEDVIAELRLQCADTRVGNMRYRGLSGGERRRVSIGVOLLNPGILLIDEP 240
QY 240 TSGLDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 299
DB 241 TSGLDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
QY 300 MVOYFAIGYPCPRYSNPADFYVDLTSIDRSRROELATREKASLAALFLKVRDLDF 359
DB 301 MVOYFAIGYPCPRYSNPADFYVDLTSIDRSRROELATREKASLAALFLKVRDLDF 360
QY 360 LMKAEKELINTSHYVSLTLQDTDC-GTAVELPGHIEQSTLIRQISNDPDLPTLLI 419
DB 361 LMKAEKELINTSHYVSLTLQDTDC-GTAVELPGHIEQSTLIRQISNDPDLPTLLI 419
QY 420 HGEACLMSTIGFLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 479
DB 420 HGEACLMSTIGFLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 479
QY 480 YELEDGLYTGPFPAKILGELPEHCAYIITGMPYVLANLRPGIOPFLHFLVWLYV 539
DB 480 YELEDGLYTGPFPAKILGELPEHCAYIITGMPYVLANLRPGIOPFLHFLVWLYV 539
QY 540 FCCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLITVPAMISKVSFLRMC 599
DB 540 FCCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLITVPAMISKVSFLRMC 599
QY 600 FEGIMAKIOFSKRTYKMPGLNLTIVANSQDKILSAMELDYPVLAIVLIGSGFVWLYV 659
DB 600 FEGIMAKIOFSKRTYKMPGLNLTIVANSQDKILSAMELDYPVLAIVLIGSGFVWLYV 659
QY 660 VSLRFIKORPSODM 673
DB 660 VSLRFIKORPSODM 673

RESULT 2

Q8C105 PRELIMINARY: PRT: 672 AA.
AC 08C105;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Sterolins 2.
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Yu H., Lu K., Lee M., Pandit B., Patel S.B.;
RT "The rat Abcg8 and Abcg9: characterization, chromosomal assignment and
genetic variation in s10sterolemic rats.";
RT Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY145899; AAN64276.1; -

SO SEQUENCE 672 AA; 75906 MW; 2FE0846E71BD9D47 CRC64;
Query Match 80.9%; Score 2835.5; DB 11; Length 672;
Best Local Similarity 79.9%; Pred. No. 2.1e-213;
Matches 538; Conservative 57; Mismatches 77; Indels 1; Gaps 1;

QY 1 MAGKAAEERGLPGKAPPDITSGLQDLRFSSSDNSLYFTYSGQPNLLEVRDLYOYDLA 60
DB 1 MAEKTEETQLMNGTVLQDASGLQDLRFSSSDNSLYFTYSGQSNLLEVRDLYOYDLA 60
QY 61 OVPWFQDLAOFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
DB 61 OVPWFQDLAOFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
QY 121 GRHGGRKSGQIMNGOPSSPOLVRKCVAVRHQNDLPLNLYRETLAIAQRLPRTF 180
DB 121 GRHGGRKSGQIMNGOPSSPOLVRKCVAVRHQNDLPLNLYRETLAIAQRLPRTF 180
QY 181 SOAQRDRVEDVIAELRLQCADTRVGNMRYRGLSGGERRRVSIGVOLLNPGILLIDEP 240
DB 181 SOAQRDRVEDVIAELRLQCADTRVGNMRYRGLSGGERRRVSIGVOLLNPGILLIDEP 240
QY 241 SGLDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
DB 241 SGLDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
QY 301 MVOYFAIGYPCPRYSNPADFYVDLTSIDRSRROELATREKASLAALFLKVRDLDF 360
DB 301 MVOYFAIGYPCPRYSNPADFYVDLTSIDRSRROELATREKASLAALFLKVRDLDF 360
QY 361 WKAETFDLEDNCEVSVTPPLDNTCLPSPFKMGAVQOFTTLIRQISNDPDLPTLLI 420
DB 361 WKAETFDLEDNCEVSVTPPLDNTCLPSPFKMGAVQOFTTLIRQISNDPDLPTLLI 420
QY 421 GAEACLMSTIGFLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 480
DB 421 GAEACLMSTIGFLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 480
QY 481 YELEDGLYTGPFPAKILGELPEHCAYIITGMPYVLANLRPGIOPFLHFLVWLYV 540
DB 481 YELEDGLYTGPFPAKILGELPEHCAYIITGMPYVLANLRPGIOPFLHFLVWLYV 540
QY 541 CCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLITVPAMISKVSFLRMC 600
DB 541 CCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLITVPAMISKVSFLRMC 600
QY 601 FEGIMAKIOFSKRTYKMPGLNLTIVANSQDKILSAMELDYPVLAIVLIGSGFVWLYV 660
DB 601 FEGIMAKIOFSKRTYKMPGLNLTIVANSQDKILSAMELDYPVLAIVLIGSGFVWLYV 660
QY 661 SLRFIKORPSODM 673
DB 661 SLRFIKORPSODM 673

RESULT 3

Q9ARU4 PRELIMINARY: PRT: 668 AA.
AC 09ARU4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Putative ABC transporter.
GN P0445D12.3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX NCBI_TaxID=4530;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsunoto T., Yamamoto K.;

Oryza sativa n1pnonbare(GA3) genomic DNA, chromosome 1, PAC clone: P0445D12.
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AP003046; BAB40032.1; -
 DR Gramene: O9AR04; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding; Transport.
 SQ SEQUENCE: 668 AA; 73368 MW; D1875B8C75B0F3B2 CRC64;
 Query Match 21.2%; Score 742; DB 10; Length 668;
 Best Local Similarity 30.3%; Pred. No. 1.8e-49;
 Matches 188; Conservative 124; Mismatches 252; Indels 56; Gaps 12;
 QY 75 WTSPSCNSCELG-----IQNLSPYVSQGMALIGSSGCGRASILDVITGR--GHGK 126
 Db 58 MARITCALKNRGDVARFLSNMSEKSKRLALMGPSGKTTLLNVLAGULTASPL 117
 QY 127 IKSQIWINQSPSSPOLVRCVAHROHQLNLTVRETIAFIAQMRTPRFSQAORDK 186
 Db 118 HLSGFLYINGRPISSEGYK--IAYVROEDLFESQLVRETLSIAELQRLTLPERKES 175
 QY 187 RVEDVIAELRLRCADTRVGMVYRGLSGERRRVSIGVOLLNPGILLDEPTSGIDSF 246
 Db 176 YVNDLFLRLGLVNCADISVDKARVIGSGEKRLSLACELASPSIIFADEPTTGIDAF 235
 QY 247 TANNVLTSLRLAKGNRLVLSLHOPRSDIFRLFDVLMTSGPTIYLG-AAQHMVQYFK 305
 Db 236 QAEKVMETLRQLAEDGHVYICSHPRGSYVKGFDIVLSEBEVYIMKPAKEPLLYFA 295
 QY 306 AIGYPCPRYSNPADFVYDLTSDRSREDEPLATREKAQSLAALFEKRVLDLFLKAE 365
 Db 296 SLGTHCPDVNPAEFLADLISVDSAESVOSRKRILENIEEFSNKV-----AT 346
 QY 366 KDLDEDTCESSVT-PLDINCLP-----SPTK-MPNAVQVFTLLIRQISNDFRDLPTL 418
 Db 347 ES-----NSSLTNDEGSEFSFKLIQKSTTKHRRMMWRFLLFRAMQAFRDGPYTK 399
 QY 419 IHGAECALMSMTIGLFYFGHSIOLSPMDTAALEFICALIPFNVLIDVISCYSERAM 478
 Db 400 VARBSVASAIIFGSYFMMKGTQISIDRMLQVLTALNTAALTKTYGCFPERKRAIV 459
 QY 479 YLEDEGLYTTGPFYFAKILGELPEHCAYIITGMPTWLANLRPGIQLPFLHFLVWLV 538
 Db 460 DBERAKGSYALGPYSSKLLAEIPGAAPLIFGSLIYPMKLPFFSFAFCGLVAYE 519
 QY 539 VECRCRIMALAAALPTFHMASFNSALNSFYLAGFPMINLSLMTVAAMSKYSFLW 598
 Db 520 SFRASMGTLTGMAATTEAAMALGPSLTATVIFGCGYVNPDPNVPVIRWIPKXSLIWM 579
 QY 599 CEEGLAKIQF-----SRTYKMPGLNLTIAVSGDKLISAMELDSPLAIIYVIGLSG 652
 Db 580 AROGICLINFKGLQFQOHSYDIQGE-----QALNFSLGGIRIADTLVAO 626
 QY 653 GEMVLYYSLRFI---KQK 669
 Db 627 GRILMFYWLTYLLTKNRP 646
 RESULT 4
 Q9L182. PRELIMINARY; PRT; 672 AA.
 AC Q9L182;
 DT 01-OCT-2000 (TRENBLREL. 15, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
 DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE ABC transporter-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; Pubmed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP001313; BAB03081.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR006162; Pplantn_attach.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
 KM ATP-binding.
 SQ SEQUENCE: 672 AA; 75269 MW; 20B2D99215600135 CRC64;
 Query Match 21.1%; Score 739.5; DB 10; Length 672;
 Best Local Similarity 30.6%; Pred. No. 2.8e-49;
 Matches 221; Conservative 126; Mismatches 251; Indels 125; Gaps 27;
 QY 7 EERGLPK--GATPQDTSGLDRLFSSES-----DN-----SLFTYSGQPN 45
 Db 7 QESSFPKTSANHHERSPVOENFSSPBYNPCDLDDNDHDSHRSQSVLRQSLRP 66
 QY 46 TLEVRDLNTQVD-----IASQVWFEQLAQFMPTSPSCNSCELGIONLSFKR 96
 Db 67 ILKFEELFYISKQSGKSGSYWFGSQEPKRNRLV-----KCVSGI-----VK 108
 QY 97 SQGMALIGSSGCGRASILDVITGRGHGKIKSQIWINQSPSSPOLVRCVAHROHNO 156
 Db 109 PGEILLAMLPSGSGKTLVTALAGRLQ-GKL-SGVSYNGEPTSSVKRK-TGFVYQDQV 165
 QY 157 LPLNLTVRETIAFIAQMRTPRFSQAORDKRVEDVIAELRLRCADTRVGMVYRGLSG 216
 Db 166 LPHLVTMETLYTALRLPKELTRKELEYEAMVSDGLTRCCNSVIGGLIRKISGG 225
 QY 217 ERRRVSIGVOLLNPGILLDEPTSGIDSFTHNLVKTLSRLAKGNRLVLSLHOPRSDI 276
 Db 226 ERRKVSIGQEMLVNPELILLDEPTSGIDSTTAARIATRLSLARGRTVVTIHOSSSL 285
 QY 277 ERLFDVYLTMTSGPTIYLGAAQHMVQYFAIGY-PCPRYSNPADFVYDLT---SIDRSR 332
 Db 286 YRFPDKVIALSECPYISGDSGVMEYFSGISGVOPSSVFNPADEVLDIANGITSDTKY 345
 QY 333 EOLATREKAQSLAALFEKRVLDLDFLKAETKDLDEDTCESSVT-PLD----- 382
 Db 346 DQ-IETNGRLDR-----LEQNSVKOSLISSTKKNLYPPLKEEVSTFEPQDPTNARLRK 399
 QY 383 --TNCLPSPTKMPGAVQVFTLLIRQIS-----NDFRDLPTLLIHGAECALMSMTIGLYE 436
 Db 400 AITNRWPTSMW-----QPSVLLKRGKRSKSHESGRLAIFWMS-----VSLSGGLWM 449
 QY 437 GHGSIOLSPMDTAALEFIPVYIIVDYSKCYSERAMLYELEDGLYTTGYFFAK 496
 Db 450 -HSRV-AHLQDQVGLFFPSIFWGFPELNAIFTEPOERPMILIKRSSGIYRLSYIAR 507
 QY 497 ILGELPEHCAYIITGMPTWLANLRPGIQLPFLHFLVWLVVVECCRIMALAAALPTF 556
 Db 508 TVGDLPMEILPFTITVTITWNGKPSLFTFIMTLMIVLVNVLVAQGLALGAILMDA 567


```

OY 557 HMAFSNALYNSFLYAGFMILSLMTVP---AMISKVSFLRMCCEGLMKIOFS-RRY 612
   568 KKAATLSSVLMFLFAGGYIQ-----HIPGFIAMLKVSFSHYCKLIVGYQYTMDEV 622
OY 613 YKAPLG-----NITFVSGDKIISAEIDSPYIATYLVIGLSGCEPMYLYV 660
   623 YEGCSLHCSVMDYEGIKMLRI---GNMMVDIAL-----AVMLL-----YRVLAYL 667
OY 661 SLR 663
   668 ALR 670
Db

RESULT 5
O92035 PRELIMINARY; PRT; 725 AA.
AC O92035:
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Putative ABC transporter.
GN AT2G01320.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Coppenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
* Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.
RT Nature 402:761-768(1999).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AC006200; AADI4532.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR Pfam: PF00005; ABC_tran. 1.
DR PRODOM: PD000006; ABC_transporter. 1.
DR SMART: SM00382; AAA. 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ
SEQUENCE 725 AA; 78899 MW; 7DB2E556FE353D7 CRC64;

Query Match 21.0%; Score 735.5; DB 10; Length 725.
Best Local Similarity 30.0%; Pred. No. 6,5e-49;
Matches 186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;

OY 75 WTPSPC-----ONSCELCIONLSFKVRSGOMAIIGSSGCRASLADYTG-----RG 122
   72 WNIITCSLSDKSSKSVFLKLVNGBAKGRLLAATMGPSGSKTLLVLAQGLSLSPRL 131
OY 123 HGGKIKSGQIWINGCPSSPOLYRKCVAHVRQINOLPLNITVETLAFLIAQMLPRTFSGA 182
   132 H-----LSGLEVNGKPRSSSKAYK--LAIFRQEDLFSSQLTVRETLISFAELQLPEISSAE 185
Db
OY 183 QNDKREVDYIAELRLROCADTRVGNMYVGLSGGERRRYSIVOLLMMNGIILDEPTSG 242
   186 EBDDEVYNNMLLKLGLVSCADSCVGNKRVGISGGEKKRLSLACELIASVIFADEPTTG 245

```

```

OY 243 LDFETAHNLVKTSLRKLAKGNRLVLSLHOPRSDFLEFLDYLMTSGTPYIYG-AAQHWY 301
   246 LDFAPQKWEIETQKLAQDGHVYCSIHOPRSGSVAKFEDYVLTGCTLYVAPACKREPL 305
OY 302 QYFATGYPCPRYNSNADYVVDLTSDIRSRREQLATREKAQIALFLEKYVDLDDFLW 361
   306 TYFGNGFLCEPHVNAEFLADLSVDYSSSEVYSSQKRVHLYVAFSGR----- 356
OY 362 KATKLDLDEDTCEVSSV---TPLDNCLDSPK-----MPGAVQDFTLLIR 405
   357 -----SSSVLYATPLS---MKETKNGKRPRAKAIYERTDGMMRQFILLKR 400
OY 406 QISNDFRDLPTLLIHAEACIMSMITGIFLFGHGSQISFNQPAALLFNGALIPRVLL 465
   401 AMQASRDGPTNKRVRMSVASAVITGVSFWRGKSQTSIQDMGLL-OVAAL---NTAM 456
OY 466 DVISKCY-----SERAMLYELEDEGLTYTGPYFPAKILGELPEHCAYIIYGMPTYIAMI 521
   457 ALFTKVGVFPKRAIVDRERSKGSYSLGPIYLSKTIATPIGAAPPLMGAVLYPMAL 516
OY 522 REGLOPPLHFLVWLWVFCRIMALAAALLPTFMASFSNALYNSFYLAGFMINUS 581
   517 NPTLSRFGRFCGIVFVSEFSAASMGITVGAMPSTEAANAAGVPSLMTVFYVGYVND 576
OY 582 SLMTVPAWISKVSFLRMCCEGLMKIOFS-----RTYKMPGKNTL---IANSGDKILSA 632
   577 NPTIFRWIPRASLRIMAFQGLCINFEGLKFDHONTFVQGEALERLSFGGRIRRET 636
OY 633 MELDSYPLY-----AIYLIY 647
   637 IAAQSRILMFWSATYLL 655
Db

RESULT 6
O9ASR9 PRELIMINARY; PRT; 725 AA.
ID O9ASR9:
AC O9ASR9:
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE At2G01320/F10A8.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN 11
RP SEQUENCE FROM N.A.
RC Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinozaki K., Ban J.,
RA Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamuya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
* Arabidopsis cDNA clones.
RT Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC Cheuk R., Chen H., Kim C.J., Shinozaki K., Ban J., Bowser L.,
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamuya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
* Arabidopsis ORF clones.
RT Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF367318; AK32905.1; -
DR EMBL: AF133617; AA091447.1; -
DR InterPro: IPR003439; ABC_transporter.

```

DR Pfam; PF00005; ABC_tran; 1.
 DR Prodom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SO SEQUENCE 725 AA; 78998 MW; 68A7E556E2EE3D7 CRC64;

Query Match 21.0%; Score 735.5; DB 10; Length 725;
 Best Local Similarity 30.0%; Pred. No. 6.5e-49;
 Matches 186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;

75 WTSPSC-----QNSCELGIONLSFKRSGOMLAITSSGCGRASLDVITG-----RC 122
 123 HGKRIKQIWMINQPSPOLVRKCAVAVRQHNOLNLTAVRETLAIAQMRIPRTSOA 182
 132 H-----LSGLEVNKRPSSSKAYK--LAFVQEDLFFSQLVRETLSPRAEDQLEISSAE 185
 183 QDRKVEDVIAELRLRQCADTRVGNMYRGLSGERRRVSIGVOLMNPGLIDDEPTSG 242
 186 ERDEYVNNLLKLTIVSCADSCVDAKVRGISGEEKRLSLACELIASPEVAFDEPTTG 245
 243 LDSTFANNVKTLSRLAKGRVLISLHQRSDIFRLFDVLMTSGTPIYLQ-AAQHMV 301
 246 LDMFAQKVMETLOKLAQDGHVYCSIHPRGVSAYAKFDIVLTETGLVYAGPAGREPL 305
 302 QYFATIGYPCPRYSNPAFYVDLTSIDRSREQLATREKROSLALFLEKVRDLDQFLM 361
 306 TYFGNGFCLPEHVNRAFLADLISDYSSSEYVSSQKRVHLDVAFSOR----- 356
 362 KAETKLDIEDYCEVSV--TPLDNCLPSPTR-----MPCAVQOFTTLIR 405
 357 -----SSVLYATPLS---MKETKNGMRPRKKAIVERIDGMWRQFELLKR 400
 406 QISNDPRDLPTLIHGAECALMSMTIGFLFGHSIQLSFMDTALLFMIGALIPENVIL 465
 401 AMQASRDGPTNVRARMSVASAVIFGVSVMKGSQTSIODRGLL-QVAAL--NTAM 456
 466 DVTSKCY-----SERAMLYELEDGLYTTGPFYFAKILGELPEHCATIIYGMPTLANL 521
 457 AALTITVGVPPKRAIVYDNRKSGSLGTYLLSKTIAETPICAAPLEMGAVLYPAPRL 516
 522 RPLGLOPFLHFLVWLVFPCRIAMALAAALLPTFMASFEFNALNSFYLACGEMINLS 581
 517 NPLTSRFGKCGIVTESFSAASMGILVGMVSTENAAVSGSLMTVFVFEFGITYNAD 576
 582 SLMTVAVMTSKVSELMCEGMLKIOPS-----RRTYKMPGLNLT--IAVSGDKILSA 632
 577 NPLIIFRWIPRASLIIMAFQGLINFEFGSLKFDHQMTFVQTEQALERLSFGGRIRER 636
 633 MELDSYPLX-----AIIIV 647
 637 IAAQSRILMFWTSATLL 655

DB 637 IAAQSRILMFWTSATLL 655

RESULT 7
 09C6M5
 ID 09C6M5 PRELIMINARY: PRT: 648 AA.

AC 09C6M5; PRELIMINARY: PRT: 648 AA.

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein (ABC transporter, putative).
 GN F27M3_2 OR ATIG31770/F27M3_2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX MEDLINE-21016719; PubMed-1130712;

DR Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feildblum T.V., Feng J.-D., Fong B., Fujit C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hulzar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Miltischer J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salazar S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 RN [12]
 RP SEQUENCE FROM N.A.
 RA Haas B.J., Volfovsky N., Town C.D., Troupkhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salberg S.L.,
 RT "Full-length messenger RNA sequences greatly improve genome
 annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [13]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troupkhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.,
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP SEQUENCE FROM N.A.
 RA Syraim-cv. Columbia;
 RC Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Natusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.,
 RT "Arabidopsis thaliana full-length cDNA.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC074360; AAC60152.1;
 DR EMBL: AY088793; AAM67104.1;
 DR EMBL: AK117530; BAC42192.1;
 DR Interpro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_transporter; 1.
 DR Prodom: PD000006; ABC_transporter; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW Hypothetical protein; 72618 MW; D52A2D2434A5B9D CRC64;
 SO SEQUENCE 648 AA; 72618 MW; D52A2D2434A5B9D CRC64;

Query Match 20.8%; Score 730.5; DB 10; Length 648;
 Best Local Similarity 30.7%; Pred. No. 1.4e-48;
 Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

9 RGLPKATPDQTSGLDRFSSSES--DNSLYFTYSGQNTLEVRDILNYQVLAQYWF 66
 19 QGLPMDSDQSKSVLAFPTTSQGLQMSY-----PLTKREEVYKVKI-----E 65
 67 QLAQFKMPWTSPSCQNSCELGIONLSFKVSSGOMLAITSSGCGRASLDVITGRHGK 126
 66 QTSQCGSKMSK-----KTLNGITGMCCPGFELMLGSGSGKTTLLSALGR--LSK 118
 127 IKSQIWMINQPSPOLVRKCAVAVRQHNOLNLTAVRETLAIAQMRIPRTSOAOR 186
 119 TFSQGMVNGPPSSGICKR--TGFVAQDDVLYHLTWETLEFTALIRLPSLTROEKAE 177
 187 RVEDVIAELRLRQCADTRVGNMYRGLSGERRRVSIGVOLMNPGLIDDEPTSGLDSE 246
 178 HYDRVIAELGLNCTSMIGPFLRGISGEEKRVSIGOMLINPGLILLDEPTSGLDSE 237
 247 TANHVKYTLRLAKGRVLISLHQRSDIFRLFDVLMTSGTPIYLQAAQHMVYFRA 306
 238 TARIYVTTIKRLASGGRTVYVTHOPSSRIYHMFDEKVVLLSEGSTPIYYGAASSAVEYFS 297

```

OY 307 IGYPCRYSNPADFYVDLTS-----IDRRSREBELATREKASLALFLEKVRDLDDEFLM 361
DB 298 L6FSTSLTVNPDLLDLNGLIPPTOKETSEOEKTVK--ETLVSAYEKN1----- 347
OY 362 KAEFDLDEDTYCVESSE-----VPLDTNCLPSPFKMPCGAOQOFTTLIRROI-SNDRDLPT 416
DB 348 --STK-LAAELCNMESHSEYETKAANKLSEQKCTTMYOTVLQKRRRRRESFVK 404
OY 417 LLIHGAECALMSMTIGFLYFGHSGIQLSFPMDTALLFMIGALIPFNVLIDVSKYSERA 476
DB 405 LRIFF---OVISAFVGLGLMMH-TPKSHIQDRTALLFFSVWGWGYPLYNVFTPOEKR 460
OY 477 MLYVELEDTYGTGYFFPAKILGELPEHCAYIIITIGMPTWLANRPGLOPFLHLLW 536
DB 461 MLINERSSGMRYLSSSYFARVNGDLPLELALPTAFVFTIYMGIGLPPDTFFLSLVL 520
OY 537 LVFECCRIAMALAAALLPTFHMAFSPSNALYNSFYLAGGFMINLSMTVP--AMISKV 593
DB 521 KSVLVAOGLGALFGLMLNINIKATTLASTVTLVFLIAGGYVO-----QIPFTYWLKYL 575
OY 594 SFLNCFEGLKMKIOPSRRTY-----KMPLGNTLIYVSGDKILSAMEL 635
DB 576 SYSYCYKLLGIOTYTDDEYECSSKVCWCRVGDPEPAIKSMGLNNLMI----DVFVWGYVL 631
OY 636 DSYPLATYLLVIGLSCGFVLYYSLSR 663
DB 632 VGYRLMA-----YMALHRYKLR 648

RESULT 8
OY 09C6R7 PRELIMINARY; PRT; 646 AA.
AC 09C6R7;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE ABC transporter, putative.
GN F566.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Becker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Utebeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
DR EMBL, AC079041; AAG50724.1;
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; AAA_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.

```

```

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding.
SQ SEQUENCE 646 AA; 72342 MW; 7A9624F82FD88A6E CRC64;
Query Match 20.6%; Score 723.5; DB 10; Length 646;
Best Local Similarity 30.6%; Pred. No. 4,8e-48;
Matches 208; Conservative 119; Mismatches 262; Indels 91; Gaps 20;

OY 22 GLQDRLFSESNSLYF--TYSQPN-----TLEVRDLYOVDLASQVPMFQLOAFKMP 74
DB 20 GLPD-MSDTQKSVLAFPTIISQPLQKSMPTIKLEYVYKTI-----EOTSQMS 71
OY 75 WTSFSCNSCELGIQNLSEFKVRSQGMALIGSSGGRASLDVITRGHGKIKSGQIWI 134
DB 72 WSKSE-----KTIINGITGVCPGEFLMGLRSGSGKXTLLSALGR--LSFTSGKMY 124
OY 135 NGQSPSPOLVKCVAVHQHNLPLNLYVRELIATIAQMRPLRPSQAKRQVEDVIAE 194
DB 125 NGQSPSGCIKRR-TGEVADVDVLYPHLVTWETLFTFALLRLPSSLTREKEHVDVIAE 183
OY 195 LRLROCADTRVGNMVRGLSGGERRRVSIQVOLLNPGILIDEPISGLDFTAHNLVKT 254
DB 184 LGINCTSMIGLPLFRGISGEEKRVISIGQEMLINPSLIDDEFTSGLDSTTAHRIYTT 243
OY 255 LSLAKGNRLVLLSLHOPRSDIFRLDVLMTSGTPIYLGAAQHMVQFYTAIGYPCRY 314
DB 244 IKRLASGGRTVYTTIHOPSSRIYHMFDRVLLSESPYYGAASAVEFSSLSGSTSLT 303
OY 315 SNPADFYVDLTS-----IDRRSREBELATREKASLALFLEKVRDLDDEFLMKAETKOLD 369
DB 304 VNPADLLDLNGLIPPTOKETSEOEKTVK--ETLVSAYEKN1-----STK-LK 350
OY 370 EDTCVESSE-----VPLDTNCLPSPFKMPCGAOQOFTTLIRROI-SNDRDLPTLIHGMA 424
DB 351 AELCNMESHSEYETKAANKLSEQKCTTMYOTVLQKRRRRRESFVKLRP---Q 407
OY 425 CIMSMTIGFLYFGHSGIQLSFPMDTALLFMIGALIPFNVLIDVSKYSERAMLYELED 484
DB 408 VISVAFGLGLMMH-TPKSHIQDRTALLFFSVWGWGYPLYNVFTPOEKRMLKKESS 466
OY 485 GLYTTGPEFFPAKILGELPEHCAYIIITIGMPTWLANRPGLOPFLHLLWLVVFCRI 544
DB 467 GMTRLSYFMARNNGDLPLELALPTAFVFTIYMGIGLPPDTFFLSLVLVSVLVAG 526
OY 545 MALAAALLPTFHMAFSPSNALYNSFYLAGGFMINLSMTVP--AMISKVSLRWQCE 601
DB 527 LGALFGLMLNINIKATTLASTVTLVFLIAGGYVO-----QIPFTYWLKTLISYCYK 581
OY 602 CLMKIOPESRRTY-----KMPLGNTLIYVSGDKILSAMELDSYPLVAI 643
DB 582 LLLGIQYDDDEYECSSKVCWCRVGDPEPAIKSMGLNNLMI----DVFVWGYVLVGYRLMA- 636
OY 644 YLYIYGLSGGFVLYYSLSR 663
DB 637 -----YMALHRYKLR 646

RESULT 9
OY 08C1Q4 PRELIMINARY; PRT; 652 AA.
AC 08C1Q4;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Sterol 1.
GN ABCG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;

```

RA Yu H., Lu K., Lee M., Pandit B., Patel S.B.:
 RT "The rat Abog5 and Abog8: characterization, chromosomal assignment and
 genetic variation in stosteroemic rats."
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY145989; AAN64275.1;
 SQ SEQUENCE 652 AA; 73372 MW; 49FE7372269299D CRC64;

Query Match 20.38; Score 713; DB 11; Length 652;
 Best Local Similarity 30.08; Pred. No. 3.2e-47;
 Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;

QY 12 PKGAT-PODTSGLDRLFFSESDNSLFTYSGQNTLEVDLYQVDSQV-PWEQOLA 69
 Db 9 PEARQPHNRGSO-----SLEKGSV--TSEARHSGV--LWFSVSNRVGPW----- 55
 QY 70 QFMPTSPSCQSCGELGI-QNLSFKVSGOMLAIIGSSCGRASLIDVTGRHGKIK 128
 Db 56 -----WNKSCQCKMKRKLKDVSLYESQTCILGSSGSKTLLDLSGRLRRGT 110
 QY 129 SGQIWINGOPSSPOLVRKCAVAVRQHNQLPNTVRETLAFIAOMRLPFTSQORDKRY 188
 Db 111 EGEVFNCGELRRDQFDCVSYLLQSDVFLSLVRETLKTYTAMAL-RSSADPYDKV 169
 QY 189 EDVIAELRQCADTVGNVAVRGLSGERRRISIGVQILMPGILIDEPISGLDSFTA 248
 Db 170 EAVLTSLSHVADOMIGNVNEGSGERRRISIAOILQDPKVMALDEPTGLDCMTA 229
 QY 249 HNLVLTSLRLAKGNRLVLTSLHOPRSDFRLPDLVLTMTSGFTIYGAQHWQYTAIG 308
 Db 230 NHIVLLVLELARNRKYIYTIHOPRSELPHEHFKIALITYGELVFCGTPEMGLGFNNCG 289
 QY 309 YPCPRYSNADPFDVLTIDRSREQLATREKAOISALF-----LEKVRDL 356
 Db 290 YCPCHSNPDEFMDLTVDTQSREREIEYKRVQMLESAFRSDCHILEIETRTIL 349
 QY 357 DDFWKAETKDLDEDCVSSVTPDLNCLPSTK-MPGAVOFTLLIRQISNDRDLP 415
 Db 350 -----KTLPM-----VPEFTKNPQMFCKLGVLLRRVTRNLRNKO 385
 QY 416 TLLIHGAELMSMTIGF-LYFGHSIQLSFMDTAALEFMIGALPENVIIDVSKCS 473
 Db 386 VYIMRYOQLMGLELIFLTVLQNMMLGAVODRGLYQLVGATPRYGMNAVLFPM 445
 QY 474 EPMALYELEDLYTTPYFAKILGELPEHCAYIIITYGPTWLANLRGLOPFLHFL 533
 Db 446 LRAVSDQESODLYOKMQLAYVLAHPFSIVATVIFSSVCTWMLGLYEVARF----- 500
 QY 534 LVMVLVFCRIMAAALLPFFHMASFFSNAL-----YNSFLYAGG 575
 Db 501 -----GYFSALMLAPHLIGELFTLVLLGMVONPNIIVSIALISGLIGSG 548
 QY 576 FMINLSLMTVPAMISKVSFLRCFEGMLKIQF 608
 Db 549 FIRNIEEMPIPLKIGTFPOKCEILVNEF 581

RESULT 10

ID 0949Y4 PRELIMINARY: PRT: 662 AA.
 AC 0949Y4:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative ABC transporter protein.
 GN F17M19.11.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.-X., Pham P.K., Banh J., Banno F., Dale J.M.,

RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Kosemura E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
 RT "Full length cDNA of gene F17M19.11 (GI:12324545)."
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY50810; AAK92745.1;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 KW ATP-binding.
 SQ SEQUENCE 662 AA; 72903 MW; CD5BC0853261B45 CRC64;

Query Match 20.28; Score 709; DB 10; Length 662;
 Best Local Similarity 31.28; Pred. No. 6.8e-47;
 Matches 216; Conservative 107; Mismatches 233; Indels 136; Gaps 23;

QY 44 PNTLEVRDNLTYQVDSQVWFEQOLAQFKAPMTSPSCQSCGELGI-QNLSFKVRS----- 97
 Db 37 PIRLKFVDQCYRKINGM-----SNDSCNIKLLGLKOKSDERTSTBERT 82
 QY 98 -----GOMLAIIGSSCGRASLIDVTGRHGKIKSGQIWINGOPSSPOLVRK 146
 Db 83 ILISGYTGMTSPGEFMAVLGSPSGSKSTLLNAVAGRILGSLNLTGKILINDGKITKQTLR 141
 QY 147 CVAVRQHNQLPNTVRETLAFIAOMRLPFTSQORDKRYEDVIAELRQCADTRNG 206
 Db 142 -GEFAQDQLPHTLVRETLVALLRPRSLTRVYKRAASVISELGTICENTVVG 200
 QY 207 NMVRLSGGERRRISIGVQILMPGILIDEPISGLDSFTAHLNLTSLRLAKG-NRLV 265
 Db 201 NTFIRISGGERKRRVSIABELLNPSLVLDERTSGLDNTAALRQTLGALHAGKTV 260
 QY 266 LISLHOPRSDFRLPDLVLTMTSGFTIYGAQHWQYTAIGPCPRYSNADPFDVLT 325
 Db 261 VTSIHQPSRVRQMEFTVLLSEGKCLFVGKGDMAVYESVGFSAFPMNADFLDLA 320
 QY 326 SIDRSREQLATREK-----AQSLALEKVRDLDFLWKAETKDLDEDCVSSVTPUD 382
 Db 321 --NGVCQDDGVTEREPNVRQTLVYAY-----DTLLAPQVK-----TCIEVSHPD 365
 QY 383 TNCLEPSTKMPGAVOFTLLI-----RQISNDRDLPNTLLIHGAELC 425
 Db 366 -NARVKTFRVNGG-GITTCIATWFSQCLILHLRLKERHESFD-----LRLFOVV 415
 QY 426 LMSMTIGLYFGHSIQLSFMDTAALEMI-----GALIPFNVLIDVSKYSRAMLVE 481
 Db 416 AASILGLMMW-HSDYR-DVHDLGLFISIFWGLVPSFNVPF-----PQERATFTE 469
 QY 482 LEDGLYTGPRYFAKILGELPEHCAYIIITYGPTWLANLRGLOPFLHFLVWLVC 541
 Db 470 RASGMYTSSYFMAVLGSLSMELVLPASFLEFTYMWVYLRGIVFELLTLVLLLYLA 529
 QY 542 CRIMLAAALLPFFHMASFFSNALSYFLGFMINSLSMTVPA--WISKVSFLW 598
 Db 530 SGGGLALGAAIMDAKKASTIYTWMLAFVLGTGYVN-----KVPSCVMWKVYSTFY 584
 QY 599 CFEGLMKIQFSRRYKMPGLNLTVASGDKTILSAELDSYPLXA-----IYLLV 647
 Db 585 CYRLVLAIOYG-----SGEILIRMLGCGSKGOGASATSGACRVEEY 629
 QY 648 IGLSG-----GFWLYVSLRFLK 666
 Db 630 IGDVGMTSVGLFLMFGYRLVLAIALRLIK 661

RESULT 11

Q9C8W6

ID 09C8W6 PRELIMINARY; PRT; 609 AA.
AC 09C8W6; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative ABC transporter.
GN F17M19.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Feng B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marfiali A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RT Nature 408:816-820(2000).
RL EMBL: AC021665; AAG52231.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding.
SQ
SEQUENCE 609 AA; 67007 MW; 65D11A874E5C0B61 CRC64;
Query Match 20.0%; Score 700; DB 10; Length 609;
Best Local Similarity 31.8%; Pred. No. 3,1e-46;
Matches 210; Conservative 103; Mismatches 225; Indels 122; Gaps 22;
QY 76 TSPSCNSCELG-I-ONLSFKVRS-----GOMLAIIGSGCGRASLDVY 118
DB 2 SNDSCNIRKLGLKPKPDETRSTERTILSCVTGMISGERMAYLPGSGSKTLNANV 61
QY 119 TGRGSGIKKSGQIMINGPSSPOLYKRCVAVRHONOLPMLVRETLAFLAONRLPRT 178
DB 62 AGRHLSGML-TKILINCKITKQTLKR-TGFVADDDLLVPLVRETLVFAALLRLPRS 119
QY 179 FSOAQRDRVEDVIAELRQCADTRVGMVYRGSGERRRVSIGVOLLMPNGLIIDE 238
DB 120 LTRVOKLAAESVISELITKCENTVNGTFRIGSGGERRVSIAMELLNPSLVLIDE 179
QY 239 PTSGIDSTFANLVKTLISRLAKG-NRLVLSLAHPRSDIFRLDLVLLMTSGTPIYLGA 297
DB 180 PTSGIDATFALRLVLTAGLHKGKTVVTSIHOPSSHVFQMFDTVLVLLSEKCLFVGK 239
QY 298 OHMVOYFAIGPCPRYSNPADFYVDLSIDRSRSEGLATREK--AQLAALFLEVR 354
DB 240 RDAAYFESVSGSPAPNPADFLDLA--NGVCTDGVTEKERKNVQTLVTAY----- 292
QY 355 DLDDFLMAETKRLDEDTCEVSSVPLDTNCLPSTKMPGCAVOQFTLLI----- 403
DB 293 ---DTLAPQVK---TCIEVSHPPQD-NARFVTRVNG--GITTLIAIWFSQLCILL 341
QY 404 -----RQISNDPRDLFTLLIHGAELMSMTIGFLYFGHGSIOLSFMDPALILMT-- 455

DB 342 HRLKERHRESFD-----LRIFQVAASTICGLMMW-HSDYR-DVADRGLGLFFSTI 392
QY 456 --GALIPENVIIDVSKSCSEANMLYELEDLYTTGPRFFAKLIGELPEHCAYIITGM 513
DB 393 FNGVLEPSFNAVPTF---POERAIFTRERASGMVTLSSYFMAHVAGLSIMELVLPASFJT 448
QY 514 PTYMLANLRPGLOPFLHFLVWLVFCCRIWMAAALLPFFHMAFFSNLINSFYLA 573
DB 449 FTYMWYLRPGVLPFLTLTVLLVYASOGIGLALGAALMDAKKASTIVVTMLAVLT 508
QY 574 GGFMINLSLWTVPA---WISKVSLRMCFCGLMKIOFSRTYKMLPLNLTVAVSGDKIL 630
DB 509 GGYYVW-----KVPSGMWMMKRVSTTFYCYRLVVAIQG-----SGEEL 548
QY 631 SAMELDSIPLVA-----ITLVIGLSG-----GPMVLYVSLRFK 666
DB 549 RMLGDSKKGASATSNAGCRFEVEEYIGVGMVTSVGLFMEFGRVILAYLRLIK 608
RESULT 12
ID 08T691 PRELIMINARY; PRT; 801 AA.
AC 08T691; 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ABC transporter ABCG1.
GN ABCG1.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of Dictyostellium."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF482380; AAL91485.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ
SEQUENCE 801 AA; 90052 MW; CCC4P036CB195A3 CRC64;
Query Match 19.8%; Score 635.5; DB 5; Length 801;
Best Local Similarity 27.9%; Pred. No. 1e-45;
Matches 187; Conservative 134; Mismatches 230; Indels 119; Gaps 20;
QY 88 IONLSFKVSGOMLAIIGSGCGRASLDVYTRGSGKIK-SGOIWIINGPSSPOLYRK 146
DB 139 LTNINGHIESGTIFALIMPSPGAGKTTLLDI---AHRINISSGMVYLNKSKSPNIRK 195
QY 147 CVAVRHONOLPMLVRETLAFLAOMLPRTFSOQRDRVEDVIAELRQCADTRVNG 206
DB 196 LCGVYTOQSDSLMPSTLVRETLNFYAKLMPDPVPLKRIQVQDIIDEMGLRCADTLVG 255
QY 207 --NMVRLSGGERRRVSIGVOLLMPNGLIIDEPTSGLDSTFANLVKTLISRLAKGRL 264
DB 256 TADNKRIRISGERRRVTISTELTGLGFSVLLIDETPSGLDSTSFYSAALKKLAKSGRT 315
QY 265 VLISLHPRSDIFRLDLVLLMTSGTPIYLGAOHMVOYFAIGPCPRYSNPADFYVDL 324
DB 316 IICTHGPRSNLYDMFDLWLLDGGONTIYKNAKALEYFNANCYHSEKKNPDAFLDL 375
QY 325 -----TSID----- 328
DB 376 INTVOEQADSDDDVDNDEEIEIGGGGGSGGAGGIEDIGISISPTMGNSAVDNINKNE 435

[illegible]

RP SEQUENCE FROM N.A.
RC STRAIN-FVB; TISSUE-Liver;
RX MEDLINE=99413474; PubMed=10485464;
RA Allen J.D., Brinkman R.F., Winhold S.J., Schinkel A.H.,
R7 "The mouse Bcrp1/Mxr/Abcg gene: amplification and overexpression in
R7 cell lines selected for resistance to topotecan, mitoxantrone, or
R7 doxorubicin.";
RL Cancer Res. 59:4237-4241(1999).
DR EMBL: AF140218; AAD54216.1; -.
DR MGD; MGI:1347061; Abcg2.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR003479; ABC_transporter.
DR InterPro; IPR006162; Pentane_attach.
DR Pfam; PF000005; ABC_tran; 1.
DR Prodom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
DR

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:14:49 ; Search time 55 Seconds
(Without alignments)
1453.188 Million cell updates/sec

Title: US-09-989-981a-8

Perfect score: 3506
Sequence: 1 MAGKAAERGLPKGATPQDT.....FMVLVYSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgnt2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
2: /cgnt2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pcp:*
3: /cgnt2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
4: /cgnt2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgnt2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgnt2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pcp:*
7: /cgnt2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
8: /cgnt2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
9: /cgnt2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
10: /cgnt2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
11: /cgnt2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
12: /cgnt2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
13: /cgnt2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
14: /cgnt2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
15: /cgnt2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
16: /cgnt2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
17: /cgnt2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
18: /cgnt2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3506	100.0	673	11	US-09-989-981a-8
2	3502	99.9	673	15	US-10-090-455-7
3	2883.5	82.2	672	11	US-09-989-981a-4
4	697	19.9	651	9	US-09-837-992-3
5	697	19.9	651	11	US-09-989-981a-6
6	697	19.9	651	15	US-10-090-455-6
7	688.5	19.6	652	9	US-09-837-992-1
8	688.5	19.6	652	11	US-09-989-981a-2
9	666	19.0	657	10	US-09-866-866a-14
10	656	18.7	663	14	US-10-108-605-245
11	642.5	18.3	655	10	US-09-981-353-35
12	640.5	18.3	655	15	US-10-120-687-61
13	640.5	18.3	655	10	US-09-961-086-1
14	638.5	18.2	655	11	US-09-866-866a-10
15	638.5	18.2	655	15	US-10-090-455-5

16	634.5	18.1	655	10	US-09-866-866a-27	Sequence 27, Appl
17	618	17.6	638	14	US-10-072-621-10	Sequence 10, Appl
18	617	17.6	674	15	US-10-090-455-4	Sequence 4, Appl
19	609.5	17.4	545	15	US-10-083-357-1335	Sequence 1335, Ap
20	573.5	16.4	627	15	US-10-090-455-8	Sequence 8, Appl
21	572.5	16.3	646	14	US-10-154-452-4	Sequence 4, Appl
22	571.5	16.3	646	15	US-10-090-455-13	Sequence 13, Appl
23	570.5	16.3	646	14	US-10-154-452-8	Sequence 8, Appl
24	567.5	16.2	646	15	US-10-072-621-9	Sequence 9, Appl
25	567.5	16.2	646	15	US-10-090-455-2	Sequence 2, Appl
26	518	14.8	604	9	US-09-745-763-197	Sequence 197, App
27	464	13.2	1564	10	US-09-801-368-244	Sequence 244, App
28	397	11.3	1501	10	US-09-801-368-346	Sequence 346, App
29	385	11.0	1511	10	US-09-801-368-250	Sequence 250, App
30	359	10.2	780	15	US-10-156-761-8698	Sequence 8698, Ap
31	336	9.6	843	15	US-09-741-669-434	Sequence 13999, A
32	252.5	7.2	894	9	US-09-815-242-11909	Sequence 434, App
33	251.5	7.2	1543	15	US-10-005-3388-8	Sequence 8, Appl
34	250.5	7.1	1617	15	US-10-005-3388-6	Sequence 6, Appl
35	246.5	7.0	331	9	US-10-090-453a-2	Sequence 11909, A
36	243.5	6.9	1617	14	US-10-156-761-12256	Sequence 12256, A
37	238.5	6.8	275	15	US-09-815-242-10053	Sequence 10053, A
38	238	6.8	352	9	US-09-971-121-4	Sequence 4, Appl
39	237.5	6.8	1594	10	US-09-971-121-2	Sequence 2, Appl
40	237.5	6.8	1642	10	US-10-156-761-10969	Sequence 10969, A
41	237	6.8	388	15	US-10-090-458-2	Sequence 2, Appl
42	235.5	6.7	1638	14	US-10-090-458-5	Sequence 5, Appl
43	235.5	6.7	1642	14	US-10-005-3388-5	Sequence 5, Appl
44	235.5	6.7	1642	15	US-10-252-819-17	Sequence 17, Appl
45	231.5	6.6	203	15		

ALIGNMENTS

RESULT 1
US-09-989-981a-8
Sequence 8, Application US/0998981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: ABCG8 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989, 981A
PRIOR FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 8
LENGTH: 673
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981a-8
Query Match 100.0%; Score 3506; DB 11; Length 673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGKAAERGLPKGATPQDTSGIQLDRLEESSENSLYFTYSGQNTLEVDLWYQVDLAS 60
|||||
Db 1 MAGKAAERGLPKGATPQDTSGIQLDRLEESSENSLYFTYSGQNTLEVDLWYQVDLAS 60
QY 61 QVWFQDLAQFKKPMWTPSCQNSCELGIONLSFKVRSQMLALIGSGGCRASILDVITG 120


```
Db 61 QVWFEDQLAQFMPPTSPSCONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
QY 121 RHGGKIKSGQIWMINGOPSSPOLYVKCAVAVHROHQLLPNLVRETLAFIAOMRLPRTFS 180
Db 121 RHGGKIKSGQIWMINGOPSSPOLYVKCAVAVHROHQLLPNLVRETLAFIAOMRLPRTFS 180
QY 181 QAOQRKRVEDVIAELRLROCACTRVGNMVRGLSGGERRRRYSIGVOLLMPGILILDEPT 240
Db 181 QAOQRKRVEDVIAELRLROCACTRVGNMVRGLSGGERRRRYSIGVOLLMPGILILDEPT 240
QY 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLVLTSGTPTIYCAAOHM 300
Db 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLVLTSGTPTIYCAAOHM 300
QY 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDFL 360
Db 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDFL 360
QY 361 WKAETKRDLEDTCVSSVTPLDNTCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
Db 361 WKAETKRDLEDTCVSSVTPLDNTCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
QY 421 GAELCLMSWTIGELFEGHSIOLSPMDTALLFMGALIPENVILDVISKYSERAMLY 480
Db 421 GAELCLMSWTIGELFEGHSIOLSPMDTALLFMGALIPENVILDVISKYSERAMLY 480
QY 481 ELEDGLYTTGPFYFAKIIIGELPEHCAYIIYGMPTMYLANLRPGLOPFLHFLVWLVF 540
Db 481 ELEDGLYTTGPFYFAKIIIGELPEHCAYIIYGMPTMYLANLRPGLOPFLHFLVWLVF 540
QY 541 CCRIMALAALALPFFHNASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMCF 600
Db 541 CCRIMALAALALPFFHNASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMCF 600
QY 601 EGLMKIORSRRTKPKPLGNLTIAVSGDKILSAMELDSTPLAIVLYIGLSGGEVWLYV 660
Db 601 EGLMKIORSRRTKPKPLGNLTIAVSGDKILSAMELDSTPLAIVLYIGLSGGEVWLYV 660
QY 661 SLRFTKORPSODM 673
Db 661 SLRFTKORPSODM 673

RESULT 2
US-10-090-455-7
; Sequence 7, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090,455
; NUMBER OF SEQ ID NOS: 2002-03-01
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-7

Query Match 99.9%; Score 3502; DB 15; Length 673;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGRAAERGLPKKATPDGTSGIDRLFFSSSDNSLFTYSGOPVTLFVRDLNLTQVDIAS 60
Db 1 MAGRAAERGLPKKATPDGTSGIDRLFFSSSDNSLFTYSGOPVTLFVRDLNLTQVDIAS 60
QY 61 QVWFEDQLAQFMPPTSPSCONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
Db 61 QVWFEDQLAQFMPPTSPSCONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
```

```
Db 61 QVWFEDQLAQFMPPTSPSCONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
QY 121 RHGGKIKSGQIWMINGOPSSPOLYVKCAVAVHROHQLLPNLVRETLAFIAOMRLPRTFS 180
Db 121 RHGGKIKSGQIWMINGOPSSPOLYVKCAVAVHROHQLLPNLVRETLAFIAOMRLPRTFS 180
QY 181 QAOQRKRVEDVIAELRLROCACTRVGNMVRGLSGGERRRRYSIGVOLLMPGILILDEPT 240
Db 181 QAOQRKRVEDVIAELRLROCACTRVGNMVRGLSGGERRRRYSIGVOLLMPGILILDEPT 240
QY 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLVLTSGTPTIYCAAOHM 300
Db 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLVLTSGTPTIYCAAOHM 300
QY 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDFL 360
Db 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDFL 360
QY 361 WKAETKRDLEDTCVSSVTPLDNTCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
Db 361 WKAETKRDLEDTCVSSVTPLDNTCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
QY 421 GAELCLMSWTIGELFEGHSIOLSPMDTALLFMGALIPENVILDVISKYSERAMLY 480
Db 421 GAELCLMSWTIGELFEGHSIOLSPMDTALLFMGALIPENVILDVISKYSERAMLY 480
QY 481 ELEDGLYTTGPFYFAKIIIGELPEHCAYIIYGMPTMYLANLRPGLOPFLHFLVWLVF 540
Db 481 ELEDGLYTTGPFYFAKIIIGELPEHCAYIIYGMPTMYLANLRPGLOPFLHFLVWLVF 540
QY 541 CCRIMALAALALPFFHNASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMCF 600
Db 541 CCRIMALAALALPFFHNASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMCF 600
QY 601 EGLMKIORSRRTKPKPLGNLTIAVSGDKILSAMELDSTPLAIVLYIGLSGGEVWLYV 660
Db 601 EGLMKIORSRRTKPKPLGNLTIAVSGDKILSAMELDSTPLAIVLYIGLSGGEVWLYV 660
QY 661 SLRFTKORPSODM 673
Db 661 SLRFTKORPSODM 673

RESULT 3
US-09-989-981A-4
; Sequence 4, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; NUMBER OF SEQ ID NOS: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

Query Match 82.2%; Score 2883.5; DB 11; Length 672;
```

Best Local Similarity 81.9%; Pred. No. 6.5e-283;
Matches 551; Conservative 52; Mismatches 69; Indels 1; Gaps 1;

```

QY 1 MAGRAAERGLPKGATPQDTSGLDRLFSSESDSLFTYSGONTELEVDLNVOLAS 60
DB 1 MAETKRETOLOMNTVLDASGLDLSFSESDSLFTYSGONTELEVDLNVOLAS 60
QY 61 QVPEFEOLOPKMPTWISPCQNSCELGIONLSFKVRSQGMALITSSGCRASLIDVITG 120
DB 61 QVPEFEOLOPKMPTWISPCQNSCELGIONLSFKVRSQGMALITSSGCRASLIDVITG 120
QY 121 RGHGKIKSGOIMVNGPSSPOLYRKCAVAHROHNLPLVRETALTAQMLPPTFS 180
DB 121 RGHGKIKSGOIMVNGPSSPOLYRKCAVAHROHNLPLVRETALTAQMLPPTFS 180
QY 181 QAORDKREVDVIAELRLROCADTRVGNMYRGSGGERRRVSIGVOLLNMPGILIDEP 240
DB 181 QAORDKREVDVIAELRLROCADTRVGNMYRGSGGERRRVSIGVOLLNMPGILIDEP 240
QY 241 SGIDSTFANHLVTLISLAKGNRLVLSLHQPBSDIFRLFDVLLMTSGTPIYLGAQOM 300
DB 241 SGIDSTFANHLVTLISLAKGNRLVLSLHQPBSDIFRLFDVLLMTSGTPIYLGAQOM 300
QY 301 VOFTFTSIGHPCPRYSNPADRYDLTSDRSRQDELATREKASLALFEKYRDDEL 360
DB 301 VOFTFTSIGHPCPRYSNPADRYDLTSDRSRQDELATREKASLALFEKYRDDEL 360
QY 361 WKAETKDEDTCESSVTPIDTNCLEPPTKMGAVOQFTLLRROISNDFRDLPTLTH 420
DB 361 WKAETKDEDTCESSVTPIDTNCLEPPTKMGAVOQFTLLRROISNDFRDLPTLTH 420
QY 421 GAACALMSMTIGLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
DB 421 GAACALMSMTIGLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
QY 481 ELEDGTYTGVPYFAATIGELPRHCAYIIYIGPPTWLANLRGLOPFLHLLVLYV 540
DB 481 ELEDGTYTGVPYFAATIGELPRHCAYIIYIGPPTWLANLRGLOPFLHLLVLYV 540
QY 541 CCRIMAAALALPFRHMASFESNALYNSFYLAGGPMNLSSIMTPAMISKSELRMCE 600
DB 541 CCRIMAAALALPFRHMASFESNALYNSFYLAGGPMNLSSIMTPAMISKSELRMCE 600
QY 601 BELMKIOESRRTYKMPGLNLTAVSGDKILISAMELDSYPLAYILVIGISGFEVLYL 660
DB 601 BELMKIOESRRTYKMPGLNLTAVSGDKILISAMELDSYPLAYILVIGISGFEVLYL 660
QY 661 SLRFTKORSQDM 673
DB 661 SLRFTKORSQDM 673
QY 660 SGLMOIOFNGHLYTQIGNFTFSILGDTWISAMDLNHLVAILVIGISGFEVLYL 659
DB 660 SGLMOIOFNGHLYTQIGNFTFSILGDTWISAMDLNHLVAILVIGISGFEVLYL 659

```

RESULT 4
US-09-837-992-3

Sequence 3, Application US/09837992
Patent No. US2002008167A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
FILE REFERENCE: 018781-00602005
CURRENT APPLICATION NUMBER: US/09/837, 992
PRIOR APPLICATION NUMBER: US 60/198, 465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204, 234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3

LENGTH: 651
TYPE: PR
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human sltosterolemia susceptibility gene (SSG)
OTHER INFORMATION: amino acid sequence
US-09-837-992-3

Query Match 19.9%; Score 697; DB 9; Length 651;
Best Local Similarity 28.9%; Pred. No. 2.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```

QY 16 TPQDTSGIDRLFSSESDSLFTYSGONTELEVDLNVOLASQVPEFEOLOAKRM 75
DB 8 TPQDSMGLQVNRSSQSLECAPAT-APEPSLILHNASYSVSHRVR-PWMD-ITSCROQ 64
QY 76 TSPSCONCELGIONLSFKVRSQGMALITSSGCRASLIDVITGR-GHGKIKSGOIM 134
DB 65 TSPSCONCELGIONLSFKVRSQGMALITSSGCRASLIDVITGR-GHGKIKSGOIM 134
QY 135 NGOPSSPOLYRKCAVAHROHNLPLVRETALTAQMLPPTFSQAORDKREVDVIAE 194
DB 116 NGRALRREDFODESYVLOSPTLSSLYRETALTAQMLPPTFSQAORDKREVDVIAE 174
QY 195 LRLROCADTRVGNMYRGSGGERRRVSIGVOLLNMPGILIDEPSTGIDSTFANHLV 254
DB 175 LSLSHVADRILGNTSLGISTGERRRVSIAAQLDOPKWLFEDEPTGLDCTANOIYV 234
QY 255 LSLHAKGNRLVLSLHQPBSDIFRLFDVLLMTSGTPIYLGAQOMVOYFTAICYCP 314
DB 235 LSLHAKGNRLVLSLHQPBSDIFRLFDVLLMTSGTPIYLGAQOMVOYFTAICYCP 314
QY 315 SNADRYDLTSDRSRQDELATREKASLALFEKYRDDEL 362
DB 295 SNADRYDLTSDRSRQDELATREKASLALFEKYRDDEL 362
QY 363 AETKDEDTCESSVTPIDTNCLEPPTKMGAVOQFTLLRROISNDFRDLPTLTH 421
DB 349 AETKDEDTCESSVTPIDTNCLEPPTKMGAVOQFTLLRROISNDFRDLPTLTH 421
QY 422 AENALMSMTIGLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
DB 391 AENALMSMTIGLYHGHSIOLSFMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
QY 476 AMLYLEGLTYTGVPYFAATIGELPRHCAYIIYIGPPTWLANLRGLOPFLHLLV 535
DB 447 AMLYLEGLTYTGVPYFAATIGELPRHCAYIIYIGPPTWLANLRGLOPFLHLLV 535
QY 536 WLVPFCRIMAAALALPFRHMASFESNALYNSFYLAGGPMNLSSIMTPAMISKSE 577
DB 500 WLVPFCRIMAAALALPFRHMASFESNALYNSFYLAGGPMNLSSIMTPAMISKSE 577
QY 578 INLSIMTPAMISKSEFLRMCCEGLMKIOFSRRTYKMPGLNLTAVS 625
DB 550 INLSIMTPAMISKSEFLRMCCEGLMKIOFSRRTYKMPGLNLTAVS 625

```

RESULT 5

US-09-989-981a-6
Sequence 6, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: ABCG5 and ABCG8: Compositions and Methods of Use
CURRENT APPLICATION NUMBER: US/09/989, 981A
PRIOR FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252, 235

;; PRIOR FILING DATE: 2000-11-20
;; PRIOR APPLICATION NUMBER: US 60/253,645
;; PRIOR FILING DATE: 2000-11-28
;; NUMBER OF SEQ ID NOS: 13
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO: 6
;; LENGTH: 651
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981a-6

Query Match 19.9%; Score 697; DB 11; Length 651;
Best Local Similarity 28.9%; Pred. No. 2.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

QY 16 TPQDTSGLDRLFSSESNSLYFTYSGQNTLEVDLNOVDLASQVPMFEOLOAFKPMW 75
DB 8 TPGSGMGLQVNNSSSLEGAPAT-APPEPHSGLIHASYSHRR-PMMD-ITSCROOM 64
QY 76 TSPSCNCELCIQMLSFVNSGOMLAIYSSGCRASLDVITGR-GHGKIKSGQIWI 134
DB 65 TROI-----LKVSLYVESGOIMCIISSGSKTLLDMSGRLGRAGTF-LGEYIV 115
QY 135 NGQPSPOLVRKCVARHNOHQLPNTLVRETIAFIAQMRLPRTFSOARORKEVDYIAE 194
DB 116 NGRALRRQFODCFYVLDSDTLSSLVRETLYTALLAI-RGNPSPFOKVEAVNAE 174
QY 195 LRLROCADTRVGNMYRGLSGGERRRVSIQVOLLNPGILIDEPSTGLDFTAHNLVKT 254
DB 175 LSLSHVADRILGNYSLSIGISGERRRVSIQVOLLNPGILIDEPSTGLDFTAHNLVKT 234
QY 255 LSLRAKGNRLVLSLHOPRSDFRLFDVLLMTSGTPIYLGAAQHMVOYFAIGYPCRY 314
DB 235 LVELARRRRIYVLTIHOPRSELPFLFDKIALISFELLPCGPAMLDFFNDGYPCEH 294
QY 315 SNPADFYDLTSDRSREOLATREKASLALE-----LEKVRDLDFFLMK 362
DB 295 SNPFDFYDLTSDRSREOLATREKASLALE-----LEKVRDLDFFLMK 348
QY 363 AETNOLDDETCVSSVPTDNLCLSPK-MPGAVOQFTLIRROISNDFDLPTLLING 421
DB 349 -----KTLPM-----VPEFKDPSGVSKGLVLRRTNRLVYRKLAVITRL 390
QY 422 AEACLSMTIGFLYFG-----HGSIQLSFMDTALLPMICALIPFNVLIDYISKYSER 475
DB 391 LQNLIMGLFLFEVLVRSNVLKGAIQ-----DRGLLQFVGATPYTGMLNAVNLFPVLR 446
QY 476 AMLYELEDGLYTTGYPFANILGELPCHAVIITIGMPTWLANLRGLOPFLHFLV 535
DB 447 AVSDQESODGLYQKQOMLAVLHVLPSVAVATMIFSSVCYWTGLHPEVARF----- 499
QY 536 WLNVFCRIMALAAALPTFHMASFS-----NALYNSFYLAG-----GFM 577
DB 500 -----GYFSALLAPHLIGELFLVLVLGIVQNPNIYNSVALLSINGVLSGFL 549
QY 578 INLSLMTVPAMISVSLRMCFLMKIOPSRRTYKMPGLNLTIVAS 625
DB 550 RNIOEMPPIPKIISYTFQKCYSELVYNEFYGLNFTGSSNSVATTN 597

RESULT 6
US-10-090-455-6
; Sequence 6, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01

;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 6
;; LENGTH: 651
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-090-455-6

Query Match 19.9%; Score 697; DB 15; Length 651;
Best Local Similarity 28.9%; Pred. No. 2.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

QY 16 TPQDTSGLDRLFSSESNSLYFTYSGQNTLEVDLNOVDLASQVPMFEOLOAFKPMW 75
DB 8 TPGSGMGLQVNNSSSLEGAPAT-APPEPHSGLIHASYSHRR-PMMD-ITSCROOM 64
QY 76 TSPSCNCELCIQMLSFVNSGOMLAIYSSGCRASLDVITGR-GHGKIKSGQIWI 134
DB 65 TROI-----LKVSLYVESGOIMCIISSGSKTLLDMSGRLGRAGTF-LGEYIV 115
QY 135 NGQPSPOLVRKCVARHNOHQLPNTLVRETIAFIAQMRLPRTFSOARORKEVDYIAE 194
DB 116 NGRALRRQFODCFYVLDSDTLSSLVRETLYTALLAI-RGNPSPFOKVEAVNAE 174
QY 195 LRLROCADTRVGNMYRGLSGGERRRVSIQVOLLNPGILIDEPSTGLDFTAHNLVKT 254
DB 175 LSLSHVADRILGNYSLSIGISGERRRVSIQVOLLNPGILIDEPSTGLDFTAHNLVKT 234
QY 255 LSLRAKGNRLVLSLHOPRSDFRLFDVLLMTSGTPIYLGAAQHMVOYFAIGYPCRY 314
DB 235 LVELARRRRIYVLTIHOPRSELPFLFDKIALISFELLPCGPAMLDFFNDGYPCEH 294
QY 315 SNPADFYDLTSDRSREOLATREKASLALE-----LEKVRDLDFFLMK 362
DB 295 SNPFDFYDLTSDRSREOLATREKASLALE-----LEKVRDLDFFLMK 348
QY 363 AETNOLDDETCVSSVPTDNLCLSPK-MPGAVOQFTLIRROISNDFDLPTLLING 421
DB 349 -----KTLPM-----VPEFKDPSGVSKGLVLRRTNRLVYRKLAVITRL 390
QY 422 AEACLSMTIGFLYFG-----HGSIQLSFMDTALLPMICALIPFNVLIDYISKYSER 475
DB 391 LQNLIMGLFLFEVLVRSNVLKGAIQ-----DRGLLQFVGATPYTGMLNAVNLFPVLR 446
QY 476 AMLYELEDGLYTTGYPFANILGELPCHAVIITIGMPTWLANLRGLOPFLHFLV 535
DB 447 AVSDQESODGLYQKQOMLAVLHVLPSVAVATMIFSSVCYWTGLHPEVARF----- 499
QY 536 WLNVFCRIMALAAALPTFHMASFS-----NALYNSFYLAG-----GFM 577
DB 500 -----GYFSALLAPHLIGELFLVLVLGIVQNPNIYNSVALLSINGVLSGFL 549
QY 578 INLSLMTVPAMISVSLRMCFLMKIOPSRRTYKMPGLNLTIVAS 625
DB 550 RNIOEMPPIPKIISYTFQKCYSELVYNEFYGLNFTGSSNSVATTN 597

RESULT 7
US-09-837-992-1
; Sequence 1, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sltoseterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837.992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18

PRIOR APPLICATION NUMBER: US 60/204,234
 PRIOR FILING DATE: 2000-05-15
 NUMBER OF SEQ ID NOS: 45
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 1
 LENGTH: 652
 TYPE: PR
 ORGANISM: Mus musculus
 FEATURE:
 OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
 OTHER INFORMATION: amino acid sequence
 US-09-837-992-1

Query Match 19.6%; Score 688.5; DB 9; Length 652;
 Best Local Similarity 28.1%; Pred. No. 1.7e-60;
 Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

45 NTLEVRDLNVOYDLASQV-PWFEQLAQFKMPTSPSCONSCELGT-QNLSFKVRSQOMLA 102
 37 HSLGVLHVSYV--SNRGGPW-----WNKSCQCKMPRLKDVSLVIESQIMC 84
 103 IIGSSGGRASLLDVTGTHGKIKSGQIMINGOPSSQVLRKCAVHRQNNQLPNT 162
 85 IIGSSSGKTTLLDALSIGRLRTGLGGEVFNVCGLRDQDFQDFSYVLQSDVFLSLT 144
 163 VRETLAFIAQMLPRTFSQADRKVEDYIAELRLQCADTRVNNYVRLSGGERRRVS 222
 145 VRETLAFIAQMLPRTFSQADRKVEDYIAELRLQCADTRVNNYVRLSGGERRRVS 203
 223 IGVOLLNPGIILDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQRSDIRFLDL 282
 204 IAAQLDQPKVMILDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQRSDIRFLDL 263
 283 VLLMTSGPIYVGAQOHVQYTAIGYPCPRSNADRYVDTLSIDRSRQELATREKA 342
 264 IAILTYGELVFCGTEPEMIGFNNCGYPCPEHSNPFDEYMDLTSVDTSREREIETRYKV 323
 343 OSIALF-----LEKVRDLDFLMAETRKDDDEPTCVSSVTPPLDTNCLPSPT 390
 324 QMLECAFKESDIYHKILENTERAYL-----KILPM---VPRKT 359
 391 K-MGAVOQFTLLIRROISNDRDLPPTLLIHGAECALMSMTGF--LYFGHSIQLSFMD 447
 360 KDPGMFGKIGVLLRRVTRNLRNKAQAVIMRLVQNLINGLFLIFLLRVQNNTLKGAQD 419
 448 TALLFMIGALIPENVIIDVSKCSEKRAMLYELEDGLYTGPFYFAKILGELPERCAY 507
 420 RVGLLYQVLGATPYGMNAVANLFPMLRAVSDQESDGLYHKQMQLAVLVHVPFSVIA 479
 508 ILYGMPTMYMLANLRPGLOPFLHFLVWLAVVFCRIMALAAALLPFIHMASPFSNML- 566
 480 TVIFSSVCYWTGLTPEVARF-----GYSAALLAHLHIGEFLLVL 522
 567 -----YNSFLAGFMINISLMTVPAMISKVSFLRCEFGMLKIOFS 609
 523 GIVQNPNTIVNSIVALLSISGLIGSGFIRNIOEMPIPKILGYTFQKRYCCELLVNNEF- 581
 610 RTTYKMPIGNLTIVNSGKILISAMELDSYPLXAI-----YLYVI 648
 582 ---YGL---NFTCGSNTSM-----NHPMCALITQGVQFIETKCPGATSRTFANFLLY 629
 649 GLSGGFVYL 657
 630 GFIPALVYL 638

RESULT 8
 US-09-989-981A-2
 Sequence 2, Application US/09989981A
 Publication No. US20030049730A1
 GENERAL INFORMATION:
 APPLICANT: Hobbs, Helen H.
 APPLICANT: Shan, Bei

APPLICANT: Barnes, Robert
 APPLICANT: Tian, Hui
 APPLICANT: Tularik Inc.
 APPLICANT: Board of Regents, The University of Texas System
 TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 FILE REFERENCE: 018781-00732005
 CURRENT APPLICATION NUMBER: US/09/989,981A
 CURRENT FILING DATE: 2002-07-23
 PRIOR APPLICATION NUMBER: US 60/252,235
 PRIOR FILING DATE: 2000-11-20
 PRIOR APPLICATION NUMBER: US 60/253,645
 PRIOR FILING DATE: 2000-11-28
 NUMBER OF SEQ ID NOS: 13
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 2
 LENGTH: 652
 TYPE: PR
 ORGANISM: Mus musculus
 FEATURE:
 OTHER INFORMATION: mouse ABCG5 (mABCG5)
 US-09-989-981A-2

Query Match 19.6%; Score 688.5; DB 11; Length 652;
 Best Local Similarity 28.1%; Pred. No. 1.7e-60;
 Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

45 NTLEVRDLNVOYDLASQV-PWFEQLAQFKMPTSPSCONSCELGT-QNLSFKVRSQOMLA 102
 37 HSLGVLHVSYV--SNRGGPW-----WNKSCQCKMPRLKDVSLVIESQIMC 84
 103 IIGSSGGRASLLDVTGTHGKIKSGQIMINGOPSSQVLRKCAVHRQNNQLPNT 162
 85 IIGSSSGKTTLLDALSIGRLRTGLGGEVFNVCGLRDQDFQDFSYVLQSDVFLSLT 144
 163 VRETLAFIAQMLPRTFSQADRKVEDYIAELRLQCADTRVNNYVRLSGGERRRVS 222
 145 VRETLAFIAQMLPRTFSQADRKVEDYIAELRLQCADTRVNNYVRLSGGERRRVS 203
 223 IGVOLLNPGIILDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQRSDIRFLDL 282
 204 IAAQLDQPKVMILDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQRSDIRFLDL 263
 283 VLLMTSGPIYVGAQOHVQYTAIGYPCPRSNADRYVDTLSIDRSRQELATREKA 342
 264 IAILTYGELVFCGTEPEMIGFNNCGYPCPEHSNPFDEYMDLTSVDTSREREIETRYKV 323
 343 OSIALF-----LEKVRDLDFLMAETRKDDDEPTCVSSVTPPLDTNCLPSPT 390
 324 QMLECAFKESDIYHKILENTERAYL-----KILPM---VPRKT 359
 391 K-MGAVOQFTLLIRROISNDRDLPPTLLIHGAECALMSMTGF--LYFGHSIQLSFMD 447
 360 KDPGMFGKIGVLLRRVTRNLRNKAQAVIMRLVQNLINGLFLIFLLRVQNNTLKGAQD 419
 448 TALLFMIGALIPENVIIDVSKCSEKRAMLYELEDGLYTGPFYFAKILGELPERCAY 507
 420 RVGLLYQVLGATPYGMNAVANLFPMLRAVSDQESDGLYHKQMQLAVLVHVPFSVIA 479
 508 ILYGMPTMYMLANLRPGLOPFLHFLVWLAVVFCRIMALAAALLPFIHMASPFSNML- 566
 480 TVIFSSVCYWTGLTPEVARF-----GYSAALLAHLHIGEFLLVL 522
 567 -----YNSFLAGFMINISLMTVPAMISKVSFLRCEFGMLKIOFS 609
 523 GIVQNPNTIVNSIVALLSISGLIGSGFIRNIOEMPIPKILGYTFQKRYCCELLVNNEF- 581
 610 RTTYKMPIGNLTIVNSGKILISAMELDSYPLXAI-----YLYVI 648
 582 ---YGL---NFTCGSNTSM-----NHPMCALITQGVQFIETKCPGATSRTFANFLLY 629
 649 GLSGGFVYL 657
 630 GFIPALVYL 638

Db 507 FMMETLMMVAVSASSMALLAAGOSVSVATLMTICVFMMIFSGLLVNLTTIASWLS 566
Oy 589 WISKVSFLRMCEGIMKIQFSRRTYKMPIGNLT-----IAVSGDKIL--SAMELDSYP 639
Db 567 WLOYSIPRYGFTALQHNHEFLGQNF--CPGLNATGNPCNYATCTGEEYLVKQIDLSRWG 625
Oy 640 LYAIIYIYIGLSGGMVLYYVSLRPIKQ 667
Db 626 LMKNHVALACMIVIFLTAYIKLLEFLK 653

RESULT 13

US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; TITLE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EPI9376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 18.3%; Score 640.5; DB 11; Length 655;
Best Local Similarity 27.2%; Pred. No. 1.3e-55;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

Oy 19 DTSGIQDLRFSSSESNLSYFTYSGOPNTLEVRDNLVYDVLASQVWFMQLQFKMPTSP 78
Db 16 NTNG-----FPATASNDLKAETEGA--VLSFHNITRYKRLKSGF-----LP----- 54
Oy 79 SCNSCELCGI-QNLSFKVRSQOMLAIIGSSCGRASLDVYTGKRGHGKIKSGQIWMINGQ 137
Db 55 -CRKPEKEILSNINGIMKPG-LNALILGPTGGKSLDYLAAARDPSGL--SGDVLINGA 111
Oy 138 PSSPOLVLRK-VAAHROHNLPLNLTVEETLAFIAOMRLPRTFSQAQRKVEDYIAELR 196
Db 112 PRANF--KCNNGYVODVWVGTLTVRENLOFSALRLATMTWHEKNERINRYIOELG 169
Oy 197 LKOCADTVNGMYVAGLSGGERRRYSIGVQLMNGIILDEPTSGDLSFTAHNLVKTLS 256
Db 170 LDKVADSVMGTQIRGVSGERKRKISIGMELTIDPSILFLDEPTGGLDSSANAVLLLK 229
Oy 257 RLAKGNRLVLSLHOPRSIDIRLFDVLMTSGPTIYGAOAHMVOYTAIGYPCPRYSN 316
Db 230 RMSKGRITIFSINHOPRYSIRKLPDSTLLASGRLMFGPQOEAALGYESAGYHCEAVNN 289
Oy 317 PADFYVDTSIDR-----SREOELATRE--KAOSIALFLKVRDL--DDELMAETK-- 366
Db 290 PADFYVDTSIDR-----SREOELATRE--KAOSIALFLKVRDL--DDELMAETK-- 366
Oy 367 -----DLDEYTVESVTPIDTNCLESPTRKMPAVOQFTLLIRQJISNDRDLPTLLI 419
Db 349 LHQSGGKKKKIYFKKISYTTSPC-----HQLRWVSKRSKKNLNGPOASIA 397
Oy 420 HGAERCLMSMTIGFLYFGHSIQLSFMDTALLFMIGALLFPNVILDYISCYG----- 473
Db 398 QIIVTVGLVIGAIYGLKNDSTGIONRAGVLFEL-----TTNCFSSVSVAVE 446

Oy 474 ----BRAMVYELDEGLTGTGPFPAKLIGE-LPEHCAYIIYGMPTVLANLRGLOPF 528
Db 447 LFVVEKRLFIHEYISGYRVSSYFLGKILSDLEPTMPLPSIFCTIYVFMLKREDAF 506
Oy 529 ILHFLVLMVVECCRIIMAAALPRTMASFFESNALYNSPYLAGGMINLSLWTPA 588
Db 507 FMMETLMMVAVSASSMALLAAGOSVSVATLMTICVFMMIFSGLLVNLTTIASWLS 566
Oy 589 WISKVSFLRMCEGIMKIQFSRRTYKMPIGNLT-----IAVSGDKIL--SAMELDSYP 639
Db 567 WLOYSIPRYGFTALQHNHEFLGQNF--CPGLNATGNPCNYATCTGEEYLVKQIDLSRWG 625
Oy 640 LYAIIYIYIGLSGGMVLYYVSLRPIKQ 667
Db 626 LMKNHVALACMIVIFLTAYIKLLEFLK 653

RESULT 14

US-09-866-866A-10
; Sequence 10, Application US/09866866A
; Patent No. US2002010224A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10

Query Match 18.2%; Score 638.5; DB 10; Length 655;
Best Local Similarity 27.9%; Pred. No. 2e-55;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

Oy 80 CONSCELGI-QNLSFKVRSQOMLAIIGSSCGRASLDVYTGKRGHGKIKSGQIWMINGQ 138
Db 55 -CRKPEKEILSNINGIMKPG-LNALILGPTGGKSLDYLAAARDPSGL--SGDVLINGA 112
Oy 139 SSPOLVLRK-VAAHROHNLPLNLTVEETLAFIAOMRLPRTFSQAQRKVEDYIAELR 197
Db 113 PRANF--KCNNGYVODVWVGTLTVRENLOFSALRLATMTWHEKNERINRYIEELG 170
Oy 198 RQCADTVNGMYVAGLSGGERRRYSIGVQLMNGIILDEPTSGDLSFTAHNLVKTLS 257
Db 171 DKVADSKVGTQIRGVSGERKRKISIGMELTIDPSILFLDEPTGGLDSSANAVLLLK 229
Oy 258 LAKGNRLVLSLHOPRSIDIRLFDVLMTSGPTIYGAOAHMVOYTAIGYPCPRYSN 317
Db 231 MSKGRITIFSINHOPRYSIRKLPDSTLLASGRLMFGPQOEAALGYESAGYHCEAVNN 289
Oy 318 ADFYVDTSIDR-----SREOELATRE--KAOSIALFLKVRDL--DDELMAETK-- 366
Db 291 ADFYVDTSIDR-----SREOELATRE--KAOSIALFLKVRDL--DDELMAETK-- 366
Oy 367 -----DLDEYTVESVTPIDTNCLESPTRKMPAVOQFTLLIRQJISNDRDLPTLLI 420
Db 350 HQLSGGKKKKIYFKKISYTTSPC-----HQLRWVSKRSKKNLNGPOASIA 398
Oy 421 HGAERCLMSMTIGFLYFGHSIQLSFMDTALLFMIGALLFPNVILDYISCYG----- 473

Mon Jul 28 09:43:31 2003

us-09-989-981a-8.rapp

Page 9

Db 399 IIVTVVLGIVIGAIYFGKLNKSTGIONKRGVLEF-----TTNOCFSSVASVEL 447

QY 474 ---ERMALTYELEDGIYTTGPYFEAKIIGE-LPEHCAYIIITYGPTYWLANLRGLOPFL 529

Db 448 FVEKKLEFHEHYTISGTVRVSYSFLKLLSDLLPMBMLPISIIFTCIAYFMIGLKRPKADAFE 507

QY 530 IAEFLWLVVFCRRIMALAAALLPLPHNASFEFSNALNSYIYLAGGFMINLSLMTYPAW 589

Db 508 VMHFTLMMVAIVASSMAALIAAGOSVAVATILMTICFVPMNITSGDLVNTLTASWLSW 567

QY 590 ISKVSFLKRCFBEGLAKIOFSRRTYKMPYSGNLT-----IIVSGDKIL--SAMELDSYPL 640

Db 568 LQVFSIPRGRFTALDHEFLGONF-CFGLNATGNPNPCYATCTGEBLYVNOGIDLSPWGL 626

QY 641 YATLYLVIGSGGFMYLVVLSRFLKO 667

Db 627 KKKHVAIVACMIVETFLIATVLLKLEFLKK 653

```

OY  530 LHEFLWLWVYECRRIMALAAALLPTPEHNAFFSNALYNSTYLAGGEMINUSLMTYPAW 569
Db  508 VMHFTLMARVYASSSNALALIAAGOSVSVATLMTICEFPMKISGGLVNLTTIASWLSW 567
OY  590 ISKVSFLRMCFEGLMKTIOFSRRTYKMPGLNLT-----IAVSGDKIL--SAMELDSTPL 640
Db  568 LQFFSIPRGFTALQNEFLGQNF-CPLGIMATGNNPCMYATCTGEBEYLAKOGIDLSWGL 626
OY  641 YALTYLYIGLSGFMWLYVSLUREFIKQ 667
Db  627 WKNHVALACMIVFIETLIAYLKLFLPK 653

```

Search completed: July 25, 2003, 17:23:59
Job time : 57 secs

```

RESULT 15
US-10-090-455-5
: Sequence 5, Application US/10090455
: Publication No. US20030027259A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hongyun
: APPLICANT: Le Bihan, Stephane
: TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
: FILE REFERENCE: 100103.406
: CURRENT APPLICATION NUMBER: US/10/090,455
: CURRENT FILING DATE: 2002-03-01
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 655
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-090-455-5

```

Query Match	18.28;	Score 638.5;	DB 15;	Length 655;
Best Local Similarity	27.98;	Pred. No. 2e-55;		
Matches 175;	Conservative 131;	Mismatches 254;	Indels 67;	Gaps 177;

```

0Y 80 CONSELGI-ONLSKRVSGOMLAIITSSGCCRAHLDVYITRGHGKTKSCQIINOP 138
Db 55 CRAPFEKELISINIMKPG-LNAILGPGGGKSSLLDVLVARKDPGSL-SCDVLINGAP 112
0Y 139 SSPOLVRCC-VAHVROHQNLLPNLVARETLEAFIAOMRLPRFESQAORBRVADYIAELRL 197
Db 113 RPANF-KCNSGVYVQDDVVMGOTLVRENTGESAALRLATYITNHNKERNIRVLEEGL 170
0Y 198 ROCADTRGNMYVRGLSGGERRRVSIGVOLMNPICILIDEPTSCDSFTAHNLVKTISR 257
Db 171 DVAOSKGTQDRIKRVSGGERRKRTSIGMELTDPISLIDEPTSLDSTAAVAILLKR 230
0Y 258 LAKGNRLVLSIHLORSDIFRLFDVYLMTSGTPYILGAOIMVYFAIGPCBRSNP 317
Db 231 MSKQRTIIFSIHQGRYSIFKLFDSLTILASRLMFHQPADALYFESAGHICAYANP 290
0Y 318 ADIFYDLTSDIR---SREOLATRE-KAOSLALELEKVRDI--DDFLMAETK--- 366
Db 291 ADFFLDIINGSTAVALNRREDFKATELIEPSKODPLEIKETLAEIYVNSFPK-ETKAEI 349
0Y 367 -----DLEDDETCVSSVTPDTONCLPSPKMPAVOQFTLLIRQIJSNDRDRLPYLLIH 420
Db 350 HQLSGGEKKKIITYPEKELSYTTSFC-----HOLRWKSKSFKNLGNPOASIAQ 398
0Y 421 GAEKCLMSMTIGLFPFGHSIOLSFMDTAALLFMIGALIPENVILDYVSKCS----- 473
Db 399 IIVTVLGLVIGAIYFGLKNDSTGIONRAGVLEFL-----TNGQFFSSVANEL 447
0Y 474 ---BRAMUYLEEDGLTTPGYEFAKTIIGE-LPEHCANIIYKMPYVLAUMLRQLODFL 529
Db 448 FVYEKLEFIHIEYISGYRVSSTYFLGKLSLDLPHRMPLPSIIFTCIYVFMGLKPPADAFF 507

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:07:19 ; Search time 26 Seconds
(without alignments)
1217.268 Million cell updates/sec

Title: US-09-989-981a-8
Perfect score: 3506
Sequence: 1 MACKAAEERGLPKGATPDPD.....FNVLYYVSLRFIKRPSQDW 673

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3502	99.9	673	1 ABG8_HUMAN	Q9H221 homo sapien
2	2873	81.9	673	1 ABG8_MOUSE	Q9DH00 mus musculu
3	2829.5	80.7	673	1 ABG8_RAT	P58428 rattus norv
4	713	20.3	652	1 ABG5_RAT	Q99092 rattus norv
5	697	19.9	651	1 ABG5_HUMAN	Q9H222 homo sapien
6	691.5	19.7	652	1 ABG5_MOUSE	Q99098 mus musculu
7	656	18.7	687	1 WHIT_DROME	Q99098 mus musculu
8	653	18.6	1294	1 YOH5_YEAST	Q08234 saccharomyc
9	640.5	18.3	655	1 ABG2_HUMAN	P10090 drosophila
10	627	17.9	695	1 WHIT_ANGGA	Q29000 homo sapien
11	623.5	17.8	679	1 WHIT_CERCA	Q17320 ceratilis c
12	621	17.7	666	1 ABG1_MOUSE	Q64343 mus musculu
13	620.5	17.7	677	1 WHIT_LOCCU	Q05360 lucilia cup
14	617	17.6	678	1 ABG1_HUMAN	Q45844 homo sapien
15	600	17.1	598	1 YPC3_CAEEL	Q11180 caenorhabdi
16	583	16.6	709	1 WHIT_ANOAL	Q16928 anopheles a
17	573.5	16.4	646	1 ABG4_HUMAN	Q25172 homo sapien
18	562.5	16.0	1049	1 ADP1_YEAST	P25371 saccharomyc
19	552	15.7	666	1 SCRT_DROME	P45843 drosophila
20	511	14.6	610	1 YOS3_CAEEL	Q09466 caenorhabdi
21	464	13.2	1564	1 PDRA_YEAST	P51533 saccharomyc
22	463.5	13.2	675	1 BROM_DROME	P12428 drosophila
23	452.5	12.9	650	1 ABG3_MOUSE	Q99081 mus musculu
24	437	12.5	668	1 BROM_PROVI	Q24739 drosophila
25	434.5	12.4	1499	1 CDR2_CANAL	P78595 candida alb
26	431	12.3	1529	1 PDRT_YEAST	P04182 saccharomyc
27	424.5	12.1	1501	1 CDR1_CANAL	P43071 candida alb
28	412	11.8	1490	1 CDR4_CANAL	O74676 candida alb
29	401	11.4	1333	1 YN99_YEAST	P53756 saccharomyc
30	397	11.3	1501	1 SNQ2_YEAST	P32568 saccharomyc
31	388.5	11.1	1501	1 CDR3_CANAL	O42690 candida alb
32	388.5	11.1	1530	1 BFR1_SCHPO	P41820 schizosacch
33	385	11.0	1511	1 PDR5_YEAST	P33302 saccharomyc

ALIGNMENTS

RESULT 1	ABG8_HUMAN	STANDARD:	PRT:	673 AA.	
34	349.5	10.0	1511	1	PDR5_YEAST
35	333	9.5	1410	1	PDR5_YEAST
36	270.5	7.7	670	1	NRTC_SYNP3
37	252.5	7.2	894	1	YH1H_ECOLI
38	251	7.2	371	1	Y40S_RHISN
39	250	7.1	1704	1	ABG3_HUMAN
40	248.5	7.1	355	1	CYSA_SYNP3
41	248	7.1	352	1	AGLK_RHIME
42	244	7.0	659	1	NRTC_SYNP7
43	241.5	6.9	332	1	SMOK_RHOSH
44	241	6.9	236	1	CYSA_CHLYU
45	238.5	6.8	344	1	CYSA_SYNP7
[2]					
RP	SEQUENCE FROM N.A., VARIANTS STIGSTEROLEMIA THR-231; GLN-263; ARG-574				
RP	AND ARG-596; AND VARIANT CYS-54;				
RP	MDLINE-20553648; PubMed-11099417;				
RP	Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,				
RP	Kwiterovich P., Shan B., Barnes R., Hobbs H.H.,				
RP	"Accumulation of dietary cholesterol in sitosterolemia caused by				
RP	mutations in adjacent Abc transporters.";				
RP	Science 290:1771-1775(2000).				
[3]					
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS STIGSTEROLEMIA				
RP	HIS-184; THR-231; GLN-263; HIS-405; PRO-501; SER-543; PRO-574;				
RP	GLU-574; ARG-574; ARG-596 AND PHE-570 DEL, AND VARIANTS HIS-19;				
RP	CYS-54; LYS-238; VAL-259; LYS-400; ARG-575 AND ALA-632.				
RP	TISSE-Liver;				
RP	MDLINE-21344600; PubMed-11452359;				
RP	Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,				
RP	Pandya A., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,				
RP	Patel S.B.;				
RP	"Two genes that map to the STSL locus cause sitosterolemia: genomic				
RP	structure and spectrum of mutations involving sterol-1 and				
RP	sterolin-2, encoded by ABCG5 and ABCG8, respectively.";				
RP	Am. J. Hum. Genet. 69:278-290(2001).				
[3]					
RP	REVIEW.				
RP	MDLINE-21474438; PubMed-11590207;				
RP	Schultz G., Langmann T., Heimerl S.;				
RP	"Role of ABCG1 and other ABCG family members in lipid metabolism.";				
RP	J. Lipid Res. 42:1513-1520(2001).				
RP	"FUNCTION: Transporter that appears to play an indispensable role				
RP	in the selective transport of the dietary cholesterol in and out				
RP	of the enterocytes and in the selective sterol excretion by the				
RP	liver into bile.				
RP	"SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to				
RP	excretion.				
RP	ABCG5 along a pathway regulating dietary-sterol absorption and				
RP	excretion.				
RP	"SUBCELLULAR LOCATION: Integral membrane protein (Probable).				
RP	"- ALTERNATIVE PRODUCTS:				
RP	Event-Alternative splicing: Named isoforms=2;				
RP	Name=1;				
RP	Isoid-Q9H221-1; Sequence=Displayed;				
RP	Name=2;				

FT	VARIANT	184	184	R -> H (in sticosterolemia).
FT			/Fttd=VAR_012252.	
FT	VARIANT	231	231	P -> T (in sticosterolemia).
FT			/Fttd=VAR_012253.	
FT	VARIANT	238	238	E -> K.
FT			/Fttd=VAR_012254.	
FT	VARIANT	259	259	A -> V.
FT			/Fttd=VAR_012255.	
FT	VARIANT	263	263	R -> Q (in sticosterolemia).
FT			/Fttd=VAR_012256.	
FT	VARIANT	400	400	T -> K.
FT			/Fttd=VAR_012257.	
FT	VARIANT	405	405	R -> H (in sticosterolemia).
FT			/Fttd=VAR_012258.	
FT	VARIANT	501	501	L -> P (in sticosterolemia).
FT			/Fttd=VAR_012259.	
FT	VARIANT	543	543	R -> S (in sticosterolemia).
FT			/Fttd=VAR_012260.	
FT	VARIANT	570	570	Missing (in sticosterolemia).
FT			/Fttd=VAR_012261.	
FT	VARIANT	572	572	L -> P (in sticosterolemia).
FT			/Fttd=VAR_012262.	
FT	VARIANT	574	574	G -> E (in sticosterolemia).
FT			/Fttd=VAR_012263.	
FT	VARIANT	574	574	G -> R (in sticosterolemia).
FT			/Fttd=VAR_012264.	
FT	VARIANT	575	575	G -> R.
FT			/Fttd=VAR_012265.	
FT	VARIANT	596	596	L -> R (in sticosterolemia).
FT			/Fttd=VAR_012266.	
FT	VARIANT	632	632	V -> A.
FT			/Fttd=VAR_012267.	
SQ	SEQUENCE	673 AA;	75678 MM;	594AFDIDCIBB50F CRC64;
<hr/>				
Query Match 99.9%: Score 3502; DB 1; Length 673;				
Best Local Similarity 99.9%: Pred. No. 8 le-256;				
Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
OY	1	MAGFAABERGLPKGATPODTSGLODRLFSSSESDNSLYFYSGGPNTLEVRDLNVOVDLAS	60	
Db	1	MAGKAAEBRGLPKGATPODTSGLODRLFSSSESDNSLYFYSGGPNTLEVRDLNVOVDLAS	60	
OY	61	QVPEFDLQAKMWTSPSCNSELGIONLSFYVRSGOMLAITGSSGGCRASLLDYTG	120	
Db	61	QVPWFEDLQAKMWTSPSCNSELGIONLSFYVRSGOMLAITGSSGGCRASLLDYTG	120	
OY	121	RGHGKTKSGGIWINGOPSSPOLVRKCVAHROHNOLLPNTVRETAFLAQMRLPTFS	180	
Db	121	RGHGKTKSGGIWINGOPSSPOLVRKCVAHROHNOLLPNTVRETAFLAQMRLPTFS	180	
OY	181	QAQRDKREVDIYAELRLROCADTVGMNVYVGLSGGERRRYSIGVOLMLNPGLILDEPT	240	
Db	181	QAQRDKREVDIYAELRLROCADTVGMNVYVGLSGGERRRYSIGVOLMLNPGLILDEPT	240	
OY	241	SGLDSPFAHNLVKTLRLKAGKNRLVYLISHOPRSDFIRLEDVLMTSGPIIYGAAQH	300	
Db	241	SGLDSPFAHNLVKTLRLKAGKNRLVYLISHOPRSDFIRLEDVLMTSGPIIYGAAQH	300	
OY	301	VQYTATGYPCCPRSNPADFYVDLTSIDRSREDELTPREKAOSLALFEKYRDLDPL	360	
Db	301	VQYTATGYPCCPRSNPADFYVDLTSIDRSREDELTPREKAOSLALFEKYRDLDPL	360	
OY	361	WKAETKDLDDETCVESSVTPLDNTCLPSPTKMPGAOVOFTLLLRQJASNFRDLPTLLIH	420	
Db	361	WKAETKDLDDETCVESSVTPLDNTCLPSPTKMPGAOVOFTLLLRQJASNFRDLPTLLIH	420	
OY	421	GAEACLMASMTIGFIYFGHSIQLSFMOTALLRNIGALLIPPNYLDIVISCYSERAMLY	480	
Db	421	GAEACLMASMTIGFIYFGHSIQLSFMOTALLRNIGALLIPPNYLDIVISCYSERAMLY	480	
OY	481	ELEGKLVTTQPPYFAKLIGELPBCATIIYYGMPTYLANRQLDQPLHLPLVYVVF	540	
Db	481	ELEGKLVTTQPPYFAKLIGELPBCATIIYYGMPTYLANRQLDQPLHLPLVYVVF	540	

 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039

```

Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
181 SOAORRKEVEDYIAELRLQCANTRNGNYRVGSGGERRRYSIGVQLMNRGILLDER 240
QY      240 TSGDSFTAHNVKTLISRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 299
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
241 TSGDSFTAHNVKTLISRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
QY      300 MVOYFTAIYPCPRNSNPADFYVDLTSIDRSREDELTAREKASIAALEKVRDLDF 359
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
301 MVOYFTAIYPCPRNSNPADFYVDLTSIDRSREDELTAREKASIAALEKVRDLDF 360
QY      360 LMKAEKXDEDETCVSSVTPIDTNCPLSPKMPGAVQOFTLLIRROISNDRFDPTLLI 419
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
361 LMKAEKXDEDETCVSSVTPIDTNCPLSPKMPGAVQOFTLLIRROISNDRFDPTLLI 419
QY      420 HGAECALMSMTIGFLYFGHGIQISFMDTALLFMGALIPNVILDIYSGSERAMLY 479
Db      420 HGESEALMSLIGFLYFGHGAQKLSFMDTALLFMGALIPNVILDIYSGKSHSESMXY 479
QY      480 YELEGITVGTGYFFAKTIGELPEHCAYIYIGMPYTLANRPGLOPFLHFLVWLV 539
Db      480 YELEGITVGTGYFFAKTIGELPEHCAYIYIGMPYTLANRPGLOPFLHFLVWLV 539
QY      540 FCCRIMATAAALPFTFHMASPESNATLYNSFYLAGEMINLSIMTVPMWISKVSFLRMC 599
Db      540 FCCRIMATAAALPFTFHMASPESNATLYNSFYLAGEMINLSIMTVPMWISKVSFLRMC 599
QY      600 FEGIMKIOFSRRTYKMPGLNLTIVASGDKILSAMELDXYPLAIYILVIGLSGGEVLY 659
Db      600 FEGIMKIOFSRRTYKMPGLNLTIVASGDKILSAMELDXYPLAIYILVIGLSGGEVLY 659
QY      660 VSLRRIKORPSQDM 673
Db      660 VSLRRIKORPSQDM 673
QY      660 LSLKLKIKOKSIQDM 673
Db      660 LSLKLKIKOKSIQDM 673

RESULT 3
ABCG8_RAT
ID      ABCG8_RAT      STANDARD:      PRT:      672 AA.
AC      P58428;
DC      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DE      ATP-binding cassette, sub-family G, member 8 (sterolin-2).
GN      ABCG8.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC      STRAIN=Sprague-Dawley;
RX      MEDLINE=21344600; PubMed=11452359;
RA      Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA      Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA      Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA      Patel S.B.;
RT      "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT      structure and spectrum of mutations involving sterolin-1 and
RT      sterolin-2, encoded by ABCG5 and ABCG8, respectively."
RL      Am. J. Hum. Genet. 69:278-290(2001).
CC      -1- FUNCTION: Transporter that appears to play an indispensable role
CC      in the selective transport of the dietary cholesterol in and out
CC      of the enterocytes and in the selective sterol excretion by the
CC      liver into bile.
CC      -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC      ABCG5 along a pathway regulating dietary-sterol absorption and
CC      excretion (by similarity).
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event-Alternative splicing; Named isoforms=2;
CC      Name=1;
CC      IsoId=P58428-1; Sequence=displayed;

```

```

CC      Name=2;
CC      IsoId=P58428-2; Sequence=VSP_000054;
CC      Note-No experimental confirmation available;
CC      -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC      SUBFAMILY.
CC      -1- CAUTION: Seems to have a defective ATP-binding region.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL: AF51785; AKR4831.1;
CC      DR      InterPro: IPR003593; AAA_Atpase.
CC      DR      InterPro: IPR003439; ABC_transporter.
CC      DR      Pfam: PF00005; ABC_tran; 1.
CC      DR      ProDom: PD000006; ABC_transporter; 1.
CC      DR      SMART: SM00382; AAA; 1.
CC      DR      PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
CC      DR      PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
CC      KW      Glycoprotein; Transmembrane; Transport; Alternative splicing.
CC      FT      DOMAIN 1 412
CC      FT      DOMAIN 413 433
CC      FT      TRANSSEM 434 446
CC      FT      DOMAIN 447 467
CC      FT      TRANSSEM 468 495
CC      FT      DOMAIN 496 516
CC      FT      TRANSSEM 517 525
CC      FT      TRANSSEM 526 546
CC      FT      DOMAIN 547 568
CC      FT      TRANSSEM 569 589
CC      FT      DOMAIN 590 628
CC      FT      TRANSSEM 629 649
CC      FT      DOMAIN 650 672
CC      FT      CARBOHYD 676
CC      FT      VARSPLIC 376
CC      SQ      SEQUENCE 672 AA; 75878 MW; 6088973F3FA4C36F CRC64;
CC      Query Match 80.7%; Score 2829.5; DB 1; Length 672;
CC      Best Local Similarity 79.8%; Pred. No. 3.2e-205;
CC      Matches 537; Conservative 57; Mismatches 78; Indels 1; Gaps 1;
QY      1 MAGKAERGLPKGKTPDTSGLDRLTSSESDNSLYFTYSGQPTLEYRDLYNVGLAS 60
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
1 MAQTKKEFTQLMNGTVLQDASSLQDSVFSSESDNSLYFTYSGQPTLEYRDLYNVGLAS 60
QY      61 QVPWFEOALQAFKMPSPSCONSCELGIONISFYKRSQOMLITGSGGGRASLDVDTG 120
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
61 QVPWFEOALQAFKMPSPSCONSCELGIONISFYKRSQOMLITGSGGGRASLDVDTG 120
QY      121 RGHGKIKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRPLPTFS 180
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
121 RGHGKIKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRPLPTFS 180
QY      181 OAOQRKREVEDYIAELRLQCANTRNGNYRVGSGGERRRYSIGVQLMNRGILLDEPT 240
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
181 OAOQRKREVEDYIAELRLQCANTRNGNYRVGSGGERRRYSIGVQLMNRGILLDEPT 240
QY      241 SGDSFTAHNVKTLISRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
241 SGDSFTAHNVKTLISRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
QY      301 VQYFTAIYPCPRNSNPADFYVDLTSIDRSREDELTAREKASIAALEKVRDLDF 360
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
301 VQYFTAIYPCPRNSNPADFYVDLTSIDRSREDELTAREKASIAALEKVRDLDF 360
QY      361 WKAETKXDEDETCVSSVTPIDTNCPLSPKMPGAVQOFTLLIRROISNDRFDPTLLI 420
Db      |||||||.....|||||.....|||||.....|||||.....|||||.....
361 WKAETKXDEDETCVSSVTPIDTNCPLSPKMPGAVQOFTLLIRROISNDRFDPTLLI 420

```

QY 421 GAELACMSMTIGELFYHGSIOLSEFMDTALLFMIGALIPFNVLIDYISKYSERAMLYX 480
 DB 420 GAELACMSMTIGELFYHGSIOLSEFMDTALLFMIGALIPFNVLIDYISKYSERAMLYX 479
 QY 481 ELEDGLTTPGTPYPAKTLGELPEHCAYIIYGMPTWLANLRPGLOPFLHFLWLVWF 540
 DB 480 ELEDGLTTPGTPYPAKTLGELPEHCAYIIYGMPTWLANLRPGLOPFLHFLWLVWF 539
 QY 541 CCRIMALAAALPTFMASFEFNALYNSFYLAGCFEMINISLMTVAANISKYSFLWCF 600
 DB 540 CCRIMALAAALPTFMASFEFNALYNSFYLAGCFEMINISLMTVAANISKYSFLWCF 599
 QY 601 EGLMKIOPSRRTYKMPGLNLTIAVSGDKILSAMELDSPYATYLYIGLSGFMVLYV 660
 DB 600 SGLMOIQFNGHIYITTOIGNLTFVYPGDAMVTAMDLSHPILYIYIGISGFLSLYL 659
 QY 661 SLAFIKOKPSQDW 673
 DB 660 SLAFIKOKPSQDW 672
 RESULT 4
 ABG5_RAT STANDARD; PRT; 652 AA.
 ID ABG5_RAT 099P67;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
 GN ABCG5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Small Intestine;
 RX MEDLINE-20578753; PubMed-11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H., Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G., Dean M., Patel S.B.;
 RT Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption.;
 RL Nat. Genet. 27:79-83(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role in the selective transport of the dietary cholesterol in and out of the enterocytes and in the selective sterol excretion by the liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to ABCG8 along a pathway regulating dietary-sterol absorption and excretion (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: AF317174; AAC53098.2;
 DR InterPro: IPR003593; AAA_ATPase.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.

FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 386 406 1 (POTENTIAL).
 FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 423 443 2 (POTENTIAL).
 FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 464 484 3 (POTENTIAL).
 FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 505 525 4 (POTENTIAL).
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 530 550 5 (POTENTIAL).
 FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 625 645 6 (POTENTIAL).
 FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
 FT NR_BIND 87 94 ATP (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 652 AA; 73342 MW; 4D42PE2BABDDAD59 CRC64;
 Query Match 20.38; Score 713; DB 1; Length 652;
 Best Local Similarity 30.08; Pred. No. 5.6e-46;
 Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;
 QY 12 PKGAT-PODTSGLQDRLESSEDSNLSYFTYSGQNTLEVRDNLQYVQLASQV-PWFQOLA 69
 DB 9 PEGARGPHNNGSGV-----SLEBSGV--TSEARHSIGV--LVNFSVSNRVGPW----- 55
 QY 70 QFKMPWTPSPSCNSCELG-I-QNLSFKYKSCOMALIISSCGRASLSDVITGRHGKIK 128
 DB 56 -----WNIKSCQCKWKDKILKDVSLYIESQOTCIILSSSGKTTLLDALSGRLRRGT 110
 QY 129 SGOIWINGQSSPOLVAKCYAHVRQHQQLPNTLVRETTAFIAQMLRPFESQAQRDRV 188
 DB 111 EGEVFNVCGLRBDQFQDCVSYLQSVFSLSTVRETLRYTAMAL-RSSADFYDKV 169
 QY 189 EDVIAELRLQCADTRGNMYVRGLSGEGRRSYIGQLAMNGIILIDPESGLDSFTA 248
 DB 170 EAVTLTSLSHVADQMGNTGNFGISGERRRSIAQLQDDPVMMLDPTTGLDCMTA 229
 QY 249 HNLVKTLSRLAKNRLVLSLHQPBSDIFPLDVLVLTSGTPIYLGAQAHMYQYFTAIG 308
 DB 230 NHIVLLVLELARNRRIYIYTHQPSLEFHFPIKAILYIGELVFCSTPEMIGFNNCG 289
 QY 309 YPCPRSNPDEVDLSDIRSRDELFARERAKOSIAALF-----LEKVRDL 356
 DB 290 YPCPRSNPDEVDLSDIRSRDELFARERAKOSIAALF-----LEKVRDL 349
 QY 357 DDFLMKAEFTKDLDEDICVESVYPLDNTCLPSPTK-MPGAVQOFTLLIRQISNDRDLR 415
 DB 350 -----KTLPM-----VPEKTKPPGKFCILGVLLRVTNRNLRNG 385
 QY 416 TLLIHGAELCMSMTIGF-LYFGHSIOLSEFMDTALLFMIGALIPFNVLIDYISKYS 473
 DB 386 VVIMRLVQNLIMGLFLIFLTLRYQNMKGAVQDRGLLYQVLCARPITYGMALNAVLFPM 445
 QY 474 ERMALTYELEDGLTTPGTPYPAKTLGELPEHCAYIIYGMPTWLANLRPGLOPFLHFL 533
 DB 446 LRAVSDQESODGILYQKQMWLAVLHALPFSIYAVIFSSVCWTGLVPEVARF----- 500
 QY 534 LVWLVECCRIMALAAALPTFMASFEFNALYNSFYLAGCFEMINISLMTVAANISKYS 575
 DB 501 -----GYSAALAPLHLPEFLTLVLGVQNPNTVNSIVALLSISGLDGS 548
 QY 576 FMINLSLMTVPAMISVFLKCFEGLMKIOP 608
 DB 549 FTRNIEMPIPLKILGYFTFOKCYCCILVYNER 581
 RESULT 5
 ABG5_HUMAN STANDARD; PRT; 651 AA.
 ID ABG5_HUMAN 09H222;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
 GN ABCG5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid:9606;
 RP [1]
 RP SEQUENCE FROM N.A., AND VARIANT GLD-604.
 RC TISSUE=Liver;
 RX MEDLINE-2053648; PubMed-11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kvitserovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 RP [2]
 RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND
 PRO-419, AND VARIANT GLD-604.
 RC TISSUE=Liver;
 RX MEDLINE-20578753; PubMed-11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srilavastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RT "Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption.";
 RL Nat. Genet. 27:79-83(2001).
 RN [3]
 RN REVIEW.
 RX MEDLINE-21474438; PubMed-11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 RN [4]
 RN VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND
 SER-550, AND VARIANT GLD-604.
 RX MEDLINE-21344600; PubMed-11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srilavastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolin-1 and
 sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 in the selective transport of the dietary cholesterol in and out
 of the enterocytes and in the selective sterol excretion by the
 liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 ABCG8 along a pathway regulating dietary-sterol absorption and
 excretion.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 in the small intestine and colon.
 CC -1- DISEASE: Defects in ABCG5 are a cause of sitosterolemia
 (HMM:102501); also known as phytosterolemia or shellfish
 sterolemia. It is a rare autosomal recessive disorder
 characterized by increased intestinal absorption of all sterols
 including cholesterol, plant and shellfish sterols, and decreased
 biliary excretion of dietary sterols into bile. Sitosterolemia
 patients have hypercholesterolemia, very high levels of plant
 sterols in the plasma, and frequently develop tendon and tubercous
 xanthomas, accelerated atherosclerosis and premature coronary
 artery disease.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF320293; AAC40003.1;
 DR EMBL: AF312715; AAC53099.1;
 DR Genew; H0NC:13886; ABCG5.
 DR MIM; 605459;
 DR MIM; 210250;
 DR GO: 0030299; P:cholesterol absorption; NAS.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;
 KW Disease mutation.
 FT DOMAIN 1 383
 FT TRANSMEM 384 404
 FT DOMAIN 405 421
 FT TRANSMEM 422 442
 FT DOMAIN 443 462
 FT TRANSMEM 463 483
 FT DOMAIN 484 503
 FT TRANSMEM 504 524
 FT DOMAIN 525 528
 FT TRANSMEM 529 549
 FT DOMAIN 550 623
 FT TRANSMEM 624 644
 FT DOMAIN 645 651
 FT NP_BIND 86 93
 FT CARBOHYD 584 584
 FT CARBOHYD 591 591
 FT VARIANT 146 146
 FT VARIANT 389 389
 FT VARIANT 419 419
 FT VARIANT 419 419
 FT VARIANT 419 419
 FT VARIANT 550 550
 FT VARIANT 604 604
 FT SEQUENCE 651 AA; 72503 MW; 950BABCB6A1536 CRC64;
 Query Match 19.9%; Score 697; DB 1; Length 651;
 Best Local Similarity 28.9%; Pred. No. 8.9e-45;
 Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;
 QY 16 TPDDTSGIADRLFESESDNSLFTYSGQDNTLEVDLNYVDLASQVWFEGLAQFKMPW 75
 DB 8 TPGGSMGLQVNRSGSSLEGAPAT-APESHISGLIHASYSVSHRRV-PWMD-ITSCROOM 64
 QY 76 TSPSCONCELCIGIONLSPKVRSGOMLAIIGSGGCRASLIDYTRG-CHGCKIRKGOIMI 134
 DB 65 TROI-----LKVSLYVESGQIMCLIGSSGSKTLLDMSSRLBAGFF-LGEYIV 115
 QY 135 NGQSSPOLVRKCAVHRQHNOLLNLTVRETFLAIAQMRPRFSSQARQKRVEDVIAE 194
 DB 116 NGRALRRQFOFDCFSYVQSPDLSLTVRETLHTALAI-RGNPSPFKKVVAVMAE 174
 QY 195 LRLRCADTRYRGNMYVRLSGGERRSVIGVOLMNPILILDEFTSGLSDFTHANLYKT 254
 DB 175 LSLSHVADRLIGNSLGSIISGERRSVIAOQLDDPVMLEDEFTSLDCTANQIYVL 234
 QY 255 LSLAKGNRLVLSIHQPRSDIFRLFDVLVLTMTSGTPIYLAAAGIMVOYFAIGPCPY 314
 DB 225 LVELARRNRRIYVLTIHQPRSLFQFLDIALISTGELIFCGTPRAEMLDFPDGCPCEH 294
 QY 315 SNPADFVYDLTISIDRRSREDELATREKAQSLAALF-----LEKVRDLDLFLMK 362

```

DB 295 SNPEFMDLTVDTSKERELETSTKRVOMIESAVKRSACHTKLNIEHMKH----- 348
OY 363 AETKDEDEVCVSSVTPDNTCLPSPTR-MPGAQOFTLLIRQISNDFRDLPTLLHG 421
DB 349 -----KLPW-----VFKRTDPSGVESKLVLLRRVRRNLRNKLAVITRL 390
OY 422 ABACLMSMTIGFLYFC-----HGSIOLEFMDTALLFMGALIPNVILDIVSKYSR 475
DB 391 LQNLINGELFLFEVLKVRSNVAKGAIQ-----DRVGLLYFGVAGPTTGLNANVLPVLR 446
OY 476 ALAYLEDEGLYTTGGYFFAKILGELPEHCAYIIITGMPTYLANRPSGLPFLHFLV 535
DB 447 AVSDDESODGLYOKQOMLALVALHVLPEFSVATMIFSSVYWTGLGHPVARR----- 499
OY 536 WLVEFCRIMAAALPTFMHASPFS-----NALVNSFLAG-----GFM 577
DB 500 -----GYFSNALLPPLHIGELFLVLGLIYQNPVINSVVALLSIAGVLGSGPL 549
OY 578 INUSSLMVTPANISKVSFLRMCPEGIMKIQFSRRYKMPDGLNLTIVAS 625
DB 550 RNIGEMPPIPKIISYTFPOKYSCEILVNAEFGINTCGSSNVSVTTN 597

RESULT 6
ABG5_MOUSE
ID ABG5_MOUSE STANDARD; PRT; 652 AA.
AC 099PE8;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Stereoloin-1).
GN ABCG5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu R., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.";
RL Nat. Genet. 27:79-83(2001).
RN [2]
RP TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=20553648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kvitrovich P., Shan B., Barnes R., Hobbs H.H.;
RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.";
RL Science 290:1771-1775(2000).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC level, in the liver.
CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF312713; AMG53097.1; -
DB MGD: MG1:1351659; Abcg5.
DB InterPro: IPR003593; AAA_ATPase.
DB InterPro: IPR003439; ABC_transporter.
DB Pfam: PF00005; ABC_tran; 1.
DB PRODOM: PD000006; ABC_transporter; 1.
DB SMART: SM00382; AAA; 1.
DB PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DB PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
DB ATP-binding; Glycoprotein; Transmembrane; Transport.
KW DOMAIN
FT TRANSMEM 386 406
FT DOMAIN 407 422
FT TRANSMEM 423 443
FT DOMAIN 444 463
FT TRANSMEM 464 484
FT DOMAIN 485 504
FT TRANSMEM 505 525
FT DOMAIN 526 529
FT TRANSMEM 530 550
FT DOMAIN 551 622
FT TRANSMEM 623 643
FT DOMAIN 644 652
FT NP_BIND 87 94
FT CARBOHYD 410 410
FT CARBOHYD 585 585
FT CARBOHYD 592 592
SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 19.7%; Score 691.5; DB 1; Length 652;
Best Local Similarity 28.6%; Pred No. 2,3e-44;
Matches 188; Conservative 129; Mismatches 241; Indels 99; Gaps 16;

OY 45 NLEVRDNLNYOVDLASOV-PWEQLAOFKMPMTSPSCONSCELGI-QNLSPFVRSQOMLA 102
DB 37 HSLGLVHYSYV--SNRQGPW-----WNISCCQKMRQILKQVSLIESQINC 84
OY 103 IIGSSGCGRASLDVITGRGCGKIKSGQIMWINGOPSSPOLVRKCVAHYRQNNLLPNT 162
DB 85 ILGSSGCKTLLDLNISGRRLTGLCEVEFVNGCELRRDQDFSVYLAGDFLSLT 144
OY 163 VRETLAFTAOHRLPTFSQAQDRKREVIENELRLRQCADRTVGMNRYRGLSGERRRVS 222
DB 145 VRETLRYTAMALACS-SADTYRKVEAVWELSHVADWIGSYNFGISGERRRVS 203
OY 223 IGVOLLNMPGILILDEPTSGDSTAHMLVYTLISLRAGNRLVLSLHQPRSDIRLPDL 282
DB 204 IAAQLAOPKVMWMDPTGIDCMTANOIVLLAELARDRVITYTIHQPSRELQCHDK 263
OY 283 VLMTSGPIYVLGAQHWOVFTALGYCPRYSNADYVDTLSIDRSRQELATRKA 342
DB 264 IAILYGLVFCGPEEMAGFEFNNCGYCPCHSNPFEDMTLSDVDSREREITYRV 323
OY 343 QSLAALFEKVRDLDDEFMKAKETDLDSDVCSSVTPDNTCLPSPTR-MPGAQOFTT 401
DB 324 QMLECAFE-----SDIYHKI-LENTIEARLYKTLPT-----VEFKRDPGMGKLV 371
OY 402 LIRQISNDFRDLPTLLIHGAELMSMTIGF--LYFGHSIOLEFMDTALLFMGALI 459
DB 372 LLRRVTRLMKRNKQAVIMRLVQNLINGELFLFYLLRVQNNILKGAVDRLVGLVYAT 431
OY 460 PPNVILDIVISCYSERAMLYLEDEGLYTTGGYFFAKILGELPEHCAYIIITGMPTYMA 519
DB 432 PYTGMLNAVNLPEMIRAVSDDESODGLYHKQOMLALVALHVLPSVIATVFSSVCYTL 491
OY 520 NLREGLOPELHFLVLMVYVCCRIMALAALPTFMHASPFSNAL----- 566
DB 492 GLIPEVARR-----GYFSNALLPPLHIGELFLVLGLIYQNPVINSI 534

```


OY 567 -----YNSFYLAGFMINLSLMTVPAMISKVSFLMCEGLKIOFSRRTYMPJLNT 621
 DB 535 VALLSTISGLISGIFRMIQEMPIPKLITGFYFORCEILVNFNE-----YGL-----NFT 587
 OY 622 IAVSGDKILAMELDSYPLVYAI-----YLVLIGISGFMVL 657
 DB 588 CGGSNMSML-----NHPMCAITGCVQFIETKTCPGATSRFTANFLILGFTPALVL 638
 RESULT 7
 WHIT. DROME STANDARD: PRT: 687 AA.
 ID WHIT. DROME STANDARD: PRT: 687 AA.
 AC P10090: 09Y3A2: 09Y333;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE White protein.
 GN W OR EG:BACN33B1.1 OR:CG2759.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RA MEDLINE=90221897; PubMed=2109311;
 RA Peppling M., Mount S.M.;
 RT *Sequence of a cDNA from the Drosophila melanogaster white gene.;
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
 RT *DNA sequence of the white locus of Drosophila melanogaster.;
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niva S.,
 RA Yamamoto D.;
 RT *Dual-tagging gene trap of novel genes in Drosophila melanogaster.;
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-U., Andrews-Ffiankoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.T., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan P.R., Bouck J., Brokstein P., Brottier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.U., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasno P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mestrovic G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT *The genome sequence of Drosophila melanogaster.;
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demillies J., Cadieu E.,
 RA Dreano S., Gloux S., Laure V., Mottier S., Gallibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papadimitrakaki G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schaefer U., Jaekle H., Bucheton A.,
 RA Belinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Saites C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT *From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.;
 RL Science 287:2220-2222(2000).
 RN [6]
 RP SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=89339145; PubMed=2503416;
 RA Teartle R.G., Belote J.M., McKewon M., Baker B.S., Howells A.J.;
 RT *Cloning and characterization of the scarlet gene of Drosophila
 RT melanogaster.;
 RL Genetics 122:595-606(1989).
 CC -1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY
 CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
 CC RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE
 CC TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF
 CC TRYPTOPHAN.
 CC -1- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 CC EMBL: X01749; CAA36038.1;
 CC EMBL: X02974; CAA26716.1;
 CC EMBL: AB028139; BAA78210.1;
 CC EMBL: AE003425; AAF45826.1;
 CC EMBL: AL133506; CAB65847.1;
 CC EMBL: X76202; CAA53795.1;
 CC PIR: S08635; FYFPM.
 CC Flybase: FBgn0003996; w.
 CC GO: GO:0004888; F:transmembrane receptor activity; NMS.
 CC GO: GO:0006727; P:chromosome biosynthesis; IMP.
 CC InterPro: IPR003593; AAA_ATPase.
 CC InterPro: IPR003439; ABC_transporter.
 CC InterPro: IPR005284; Pigment_permease.
 CC Pfam: PF00005; ABC_tran.1.
 CC ProDom: PD000006; ABC_transporter.1.
 CC SMART: SM00382; AAA.1.
 CC TIGRFAMS: TIGR00955; 3a01204.1.
 CC PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 CC PROSITE: PS00893; ABC_TRANSPORTER_2; 1.

KM Pigment: ATP-binding; Transmembrane; Transport.
 FT NP_BIND 130 137 ATP (BY SIMILARITY).
 FT TRANSMEM 435 453 POTENTIAL.
 FT TRANSMEM 465 485 POTENTIAL.
 FT TRANSMEM 515 533 POTENTIAL.
 FT TRANSMEM 542 563 POTENTIAL.
 FT TRANSMEM 576 594 POTENTIAL.
 FT TRANSMEM 659 678 POTENTIAL.
 FT CONFLICT 25 29 GDSGA -> LIFEIPHYCHRYAD (IN REF. 2 AND 3).
 FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).
 FT CONFLICT 335 371 VGACCPNPNADYOVYVAVVPGREIESRDIKAC -> IRLINSYPAWVPSVPTTIRPTFTYKCMPLCPGRCSPVI GSPPYG (IN REF. 3).
 SQ SEQUENCE 687 AA; 75672 MW; 24AFAD799DED036 CRC64;
 Query Match 18.78; Score 656; DB 1; Length 687;
 Best Local Similarity 30.3%; Pred. No. 1.2e-41;
 Matches 178; Conservative 113; Mismatches 265; Indels 32; Gaps 10;
 OY 88 IONLSFKVSGOMLAIIGSSGGRASLDVITGRGHGG--KISGQIWMNGSPSPOLVR 145
 DB 113 LKAVCGVAYPGLLAWGSSGAKRTLLNALFRSPGIGVSPSGMRLNGOPADAKEMQ 172
 OY 146 KCVAHVROHNLPLNLTVEETLAFIAQMRIPFESQQRDRVEDVIAELRLQCADTRV 205
 DB 173 ARCAVYQODDLFGSLTAREHLIFQANVRPRHLITQRARVDQVIELSKCHTTI 232
 OY 206 G-NMYVRLSGGRRRVSIGVOLLNMGILLDEPTSGDSTAHNLVKTLSRLKGNRL 264
 DB 233 GVBGRVKGSGGGRKRRLAFSEALTPDLICDEPTSGDSTAHNSVYQVTKLSGKGT 292
 OY 265 VLSLQPRSDIRFLDVLVLTMSGTPPIYLGAAQHMVQYTAIGYPCPRNSPADRYVDL 324
 DB 293 VITTIQPSSELELTDKILMAEGRAVAFITPSEADFSYGAQCPITVYNADFEVQV 352
 OY 325 TSIDRRSRQELATREKAOSLAALF-LEKY-RDLDFLMAETKDLDEDECVSSVTPV 382
 DB 353 LAY---VPGREIESRRIKICDNFALSKVAROMEDL---ATKNLEK-----PLE 397
 OY 383 TNCLEPSP---TKMPGAVOOFTLLIRQISNDRFLPTLLIHGAECNLSMTIGTFYFGH 438
 DB 398 ---OPENGYTKATWFOFRVLRWSMLSVLKEPLVVRLLQTTVAIILGLLEFGQ 452
 OY 439 GSISLQSFMDTALFEMGALIPPNVLDVYSKYSRAMLYTELEGLTYTGGYFPAKIL 498
 DB 453 QLTQGVMMINGAIFLELTNMTQNFATINVTSELPEVPMRARSRLTRCDYFFLGKTI 512
 OY 499 GELPEHCAYIIYGMPTWLANLRPGLPFLHFLVWLVVFCRIMALAAALLPTEHM 558
 DB 513 AELPELTVPLVFTALAYMIGRACVLFENCLALVTVANVSTFGYILISCASSSTMS 572
 OY 559 ASFEENALYNSETLAGCFMINISLTVPAWISKVSFLRMCDFGLMKTIQFS---RRTYKA 615
 DB 573 ALSVGPVLIIPLELFGFELNSGVPVYKMLSYLSMFRANSGLLINQADVEPEEISC 632
 OY 616 PLGNLTINVSQDKIISAMELSDSPYLAIVLIVIGLSGGVAVLYVSLR 663
 DB 633 TSSNTCPSSGKVIETLNFNSADPLDVLVGLAIIIVSFRVLAIALR 680
 RESULT 8
 YOH5_YEAST STANDARD; PRT; 1294 AA.
 AC 008234; 008233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.

OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97321807; PubMed=9178509;
 RX Tzeremia M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast chromosome XV reveals eight known genes and ten new open reading
 frames including homologues of ABC transporters, inositol
 phosphatases and human expressed sequence tags";
 RL Yeast 13:583-589(1997).
 CC 1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 DR EMBL: Z74817; CA99085.1; -
 DR EMBL: Z74816; CA99084.1; -
 DR PIR: S77690; S77690.
 DR SGD: S0005435; YOL075C.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Prodom: PD000006; ABC_transporter; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 2.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 2.
 DR Hypothetical protein: ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 605 625 POTENTIAL.
 FT TRANSMEM 1039 1059 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1267 1287 POTENTIAL.
 FT NP_BIND 62 69 ATP (POTENTIAL).
 FT NP_BIND 727 734 ATP (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 983 983 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1294 AA; 145157 MW; C555500A5B5284E CRC64;
 Query Match 18.6%; Score 653; DB 1; Length 1294;
 Best Local Similarity 30.1%; Pred. No. 4.4e-41;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;
 OY 88 IONLSFKVSGOMLAIIGSSGGRASLDVITGRGHGKRSQI----- 132
 DB 45 VNTFSMDLPSSGSMAMVGGSGKTLVLNVLASKISGGLHNSIRYVLEDGSEPNETE 104
 OY 133 ---WINGPSSPOLVRKCVANRQHNQLPLNLTVEETLAFIAQMRIPFESQQRDR- 187
 DB 105 PKRAHLDGQ-DHPQKHVIMATVLPQDVLSPRLTCHETLKFADLKL-----NSSERKRL 159
 OY 188 -VEDVIAELRLQCADTRVGNMYVRLSGGRRRVSIGVOLLNMGILLDEPTSGIDSP 246
 DB 160 WVEOLIEELIKKCACTLVGDNSHRLSGGGRKRRLSIGVOMISNPSIMFLDEPTGLDAV 219
 OY 247 TANHNVKTLRLAK-CNRLVLSLHOPRSDIRFLDVLVLTMSGTPPIYLGAAQHMVQYTP 305

```

DB 220 SAFVYIKTLKLAKEDGRTFIMSIHQPRSDILFLDQVCILSKGNVVCCKMDNTIPE 279
OY 306 AIGYPCPRYSNPADFEVDITSDRSREBELTREKAOSIALFEKVDDLDMKAET 365
DB 280 SIGHYVPLVNPADFEVDITSDRSREBELTREKAOSIALFEKVDDLDMKAET 365
OY 366 KILDEDTYCVESSTPLDITNCLSPPTMGCAVOOFTLLIRQISNDPRDLPTLHGAEC 425
DB 331 ILOLOAESYI-SNATEIQIOM--TTRLP-FMKOVYVLTIRNFKLNSDVTTLSTFAEC 386
OY 426 LMSMTGIFGHSIGLSFMDTAAALFNGALIP--FNVIIDVSKCYSEBAMLYEEL 483
DB 387 IIGTCVGMWYKPKSSIGGLRTTACLYASTILOCYVILPDTYRLCQDIALDREKA 446
OY 484 DELVYTGPEFA-KILGELPEHCATIIYIGMPTYLAMRPGLOPFLHFLVLTWVCC 542
DB 447 BESVPLAFYVARKISLFLSDFAMTMFVSTIYFEGLEADARFFYFAVAVLQOLSC 506
OY 543 RIMAAALALPTFHAFSPFNALYNSFLAGFPMINLSLMTVPAMISKVSFLRCEFG 602
DB 507 SGLSMHLSVAVSRDFSKASLGVMTFTVLSMGGCFVYVNAKRVYVYVIAFTWISFET 566
OY 603 LMKIOPSR--TYKMPGLNLTAVSG 626
DB 567 LMSSTFNSYCTTDNLECLGNOILEVYG 595

RESULT 9
ABG2_HUMAN
ID ABG2_HUMAN STANDARD; PRT; 655 AA.
AC 09JUN00; 095374; 09BY73; 09NUS0.
DT 16-OCT-2001 (Rel. 40. Created)
DT 28-FEB-2003 (Rel. 41. Last sequence update)
DE 28-FEB-2003 (Rel. 41. Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-
GN binding cassette transporter) (Breast cancer resistance protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-99065313; PubMed-9850061;
RA Allkmetts R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCG2) on
RL Chromosome 4q22 that is involved in multidrug resistance.";
RN Cancer Res. 58:5337-5339(1998).
RC [12]
RP SEQUENCE FROM N.A.
RX TISSUE-Breast cancer;
RX MEDLINE-99080071; PubMed-9861027;
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
RA Ross D.D.;
RT "A multidrug resistance transporter from human MCF-7 breast cancer
RL cells.";
RN Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
RN [13]
RP ERRATUM.
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
RA Ross D.D.;
RN Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
RN [14]
RP SEQUENCE FROM N.A.
RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,
RA Sugimoto Y.;
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a
RL homodimer.";
RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RP [15]
RC SEQUENCE OF 198-655 FROM N.A.
RC TISSUE-Placenta;

```

```

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murauchi K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Makatsutsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niimiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RN Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP REVIEW.
RX MEDLINE-21474438; PubMed-11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RN J. Lipid Res. 42:1513-1520(2001).
CC -1- FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
CC IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
CC CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME
CC RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY
CC DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
CC MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF103796; AAC09188.1; -
DR EMBL: AF098951; AAC97367.1; -
DR EMBL: AB056867; BAB39212.1; -
DR EMBL: AK002040; BAA92050.1; -
DR GeneW: HGNC:74; ABCG2.
DR MIM: 603756; -
DR GO: GO:0016021; C:Integral to membrane; TAS.
DR GO: GO:0005524; F:ATP binding activity; TAS.
DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter actl. . .; TAS.
DR GO: GO:0005215; F:transporter activity; TAS.
DR GO: GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
DR GO: GO:0009315; P:drug resistance; TAS.
DR GO: GO:0006832; P:small molecule transport; TAS.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 395
FT TRANSMEM 396 416
FT POTENTIAL 417 428
FT TRANSMEM 429 449
FT DOMAIN 450 477
FT TRANSMEM 478 498
FT POTENTIAL 499 506
FT TRANSMEM 507 527
FT POTENTIAL 528 535
FT TRANSMEM 536 556
FT DOMAIN 557 630
FT TRANSMEM 631 651
FT DOMAIN 652 655
FT NP_BIND 80 87
FT CARBOHYD 418 418
FT CARBOHYD 557 557
FT CARBOHYD 596 596
FT CONFLICT 24 24
FT CONFLICT 166 166

```

E -> Q (IN REF. 2 AND 4).

```

FT CONFLICT 208 208 F -> S (IN REF. 1)
FT CONFLICT 315 316 MISSING (IN REF. 5)
FT CONFLICT 482 482 R -> T (IN REF. 2)
SQ SEQUENCE 655 AA: 72343 MW: 8946D3511DC5CE0 CRC64:

Query Match
Best Local Similarity: 27.9%; Pred. No. 1.6e-40;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

OY 80 CONSELGI-QNLKRVSGOMLAIGSSGGRASLVYITGRGKIGKQIWINQCP 138
DB 55 CRKPEKEILSNINIKMPG-LMALIGPTGGKSSLLVLAARKDPSGL-SGDVLINAP 112
OY 139 SSPOLVRKC-VAHROHQLPLNTVRETIAFIAQMRPLPFESQORRVEDYAEURL 197
DB 113 RPANP-KCNSGYVQDDVVGTLVRNLOPFAALATWTNHEKMERINRVEELGL 170
OY 198 RQCADTRVGNMYVGLSGERRRVSIGVOLLMPGILLDEPTSGDSFTAHNLVKTLSR 257
DB 171 DKVADSKVGTQPIGVSQGERKRTSIGMELITDPSILFLDEPTGIDSTANAVLLTKR 230
OY 258 LAKGNRLVLSLHOPRSDIERLDVLMTSGTPIYGAQAHMVOYFAIGPCPRYSNP 317
DB 231 HSKGRITLISIHOPRSIFKLPDSLTLLASGRMFHGPQAEALGYFESAGYHEAYNNP 290
OY 318 ADPVYDLTSLDR-SREQLATRE--KAQSLAALFLEKVRDL--DDFLMKATK--- 366
DB 291 ADPLDILINDSTVALNREDPKATELIEPSKODKPLEIAETIYVNSSTFK-ETKAEL 349
OY 367 -----DLDEPTCVESVTPIDNCLPSPTKPGAVOQFTLLIRROISNDRDLPTLII 420
DB 350 HOLSGGKRRKITYFKKISYTSFC-----HQLRVNSRSEFNLLGNQASIAQ 398
OY 421 GAELCKMTGILYFEGHSIQLSFMDTALFPIGALIPNVLDIYSKYS----- 473
DB 399 IYTVVGLVIGALYFGLKNDSTGIONRAGLFL-----TTNQCFSVSVAEL 447
OY 474 ---ERAMLYLEEDGLTYTGPYFAKILGE-LPEHCAYIIITGMPTWLANRGLDPL 529
DB 448 FVEEKKLFIEHYISGYRVSSYFGLKLSLDPHMLPSITFCITVFMGLKRADEF 507
OY 530 LHELVLVLFVCCRIMALAAALPTFHMAFSFNALYNSEYLAGCFMILNLSMTYDAM 589
DB 508 VMETLMAVAVSASSMALATAAGSVSVATILMTICFVEMMISGLLVNTITIASMLSW 567
OY 590 ISKVSFLRMCFEGMLKIQFSRRTYKMLGULT-----IAVSDKIL--SAMELSYPL 640
DB 568 LQYFSTIPRYGFTALQNHFLQONF-CPLNATGNNPCNVACTGEYLVKQIDLSPMGL 626
OY 641 YATYLVIGLGGFMVLYVSLRPIKO 667
DB 627 WKNHVALACMIVIFLTITAIKLFLKK 653

RESULT 10
WHIT ANOGA STANDARD: PRT: 695 AA.
AC 027256: 017006:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE white protein.
GN M.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-Snakoto / G3;
RX MEDLINE=96423158; PubMed=8825759;
RA Besansky N.J., Bedell J.A., Benedict M.O., Mukabayire O., Hliffaker D.,
RA Collins F.H.;

```

```

FT Cloning and characterization of the white gene from Anopheles
FT gambiae.
RL Nucleic Acids Res. 1995; 23(19):5111-5116.
CC -I- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: U29486; AAC46995.1; -
DB EMBL: U29485; AAC46994.1; -
DB EMBL: U29484; AAC47423.1; -
DB InterPro: IPR003593; AAA_Atpase.
DB InterPro: IPR003439; ABC_transporter.
DB InterPro: IPR005284; Pigment_permease.
DB Pfam: PF00005; ABC_tran. 1.
DB ProDom: PD000006; ABC_transporter. 1.
DB SMART: SM00382; AAA. 1.
DB TIGRfam: TIGR00955; 3a01204. 1.
DB PROSITE: PS00211; ABC_TRANSPORTER. 1;
DB PROSITE: PS00893; ABC_TRANSPORTER. 2; 1.
DB Pigment: ATP-binding; Transmembrane; transport.
DB NP_BIND 133 140
DB NP_BIND 288 295
DB TRANSMEM 444 464
DB TRANSMEM 474 494
DB TRANSMEM 524 544
DB TRANSMEM 552 572
DB TRANSMEM 581 601
DB TRANSMEM 669 689
DB CARBOHYD 472 472
DB CARBOHYD 645 645
DB CONFLICT 100 100
DB CONFLICT 691 693
SQ SEQUENCE 695 AA: 77218 MW: EB8B9517239B2961 CRC64:

Query Match
Best Local Similarity: 17.9%; Score 627; DB 1; Length 695;
Matches 189; Conservative 128; Mismatches 289; Indels 112; Gaps 17;

OY 14 GATPQDPSGLDRLFFSSS-----DNSLYFTYSGQPNLT-EVRDLNVOYDLASQVPEF 66
DB 10 GDAESKTYTSSRRYSSTSSYODSDMDALNTTLNDRATLQVWPKPSYSGVKQIPCE 69
OY 67 QLAQFKMPW-----TSPSC--QNSCELG-----IONLSFKVRS 98
DB 70 RLT---YTWKEIDYVFGCAPPDGKREPLCTRLRNCCYRORRDNPRKHLNVAAGVASK 126
OY 99 QMLAIGSSGGRASLDVITGRGHC-KIKSGOI-WINQPSPPQVLRKCAVAHQHNO 156
DB 127 ELLAVMGSSGAGKTTLLNALAFRSPGVKISPNVRLNGLVNAEDLRACAVQDD 186
OY 157 LRLPLVTRRETLAFIAQMRPLPFESQORRVEDYAEURLRQCADTRVGNM-YVRGLSG 215
DB 187 FIPSLTRREHLDPAMLRMGDRDPAVKYORHVOEVLQELSLVKCADITIGAPRKIGLSG 246
OY 216 GERRRVISIGVOLLMPGILLDEPTSGDSFTAHNLVKTLSRLAKGNRLVLSHQRSD 275
DB 247 GERKRLAFASSETLDPHLLCDEPTSGDSFMAHSVADVLKGAAMKRTITLTIHOPSS 306
OY 276 IFRLEFDVLMTSGTPIYGAQAHMVOYFAIGPCPRYSNPADFYDLTSLDRSREO 335
DB 307 LYGCFDRLILVAGSRVAFSLSPQSAEFFSGDLIPCPRNPNPADFYQMLAIAPAK---- 362
OY 336 LATREKQSLAALFLEKVRDLDFLWKAETKDLDEPTCVESVTPIDTNCLPSPPT----- 390

```

```

Db 363 -----EACRDMIRKICDSFVSPAREVLETAASVAGKG 396
OY 391 -KMPGAVQ-----OFTLLIRROISNDFRDLPTLLIHGAELMTMTGCF 433
Db 397 MDEPYMLOQVEGVSTGYRSSWMTQFYCLIMRSMLSYLKDPMLVAVRLLOTAAMAVTLIGS 456
OY 434 LYFGHGSIQLSFMDTAALLFMIGALIPFNVIIDVTSKCYSERAMLYELEDGLYTTGPYE 493
Db 457 IYFGVLDQDDGVNMINGSLFELTNMTFQNVFAVINVSAPLEPVLFRKRSRLYVVDYTF 516
OY 494 FAKILGELPEHCAYIIITGMPYMWLANRPGLOPPLFLVLMVAVFCCRIMAAAL 553
Db 517 LKRTIAELPLTIAPEVFTSTYPMIGLRTGATHTLTTLVTVLVANVSTSGYLISCAS 576
OY 554 PTFHMASFFSNALYNSFYLAGGFMINLSLMTVPAMIKVSELR--CFEGIMAKTQFS- 609
Db 577 SSISMAISVCPPIVPIELIFGGEFLNSAS--VPYFVYISLISFRYANALLINOMST 633
OY 610 -----RRTYKMPGLNLTIAVSGDKILSAMELDYPYAIYLIYVIGSGFMYLY 658
Db 634 VVDEIACTRANWTCPSREILLETENFRV-EDFALDIACLFR--LIVIFRLGALICIM 688

```

RESULT 11

```

WHIT_CERCA
ID WHIT_CERCA STANDARD: PRT: 679 AA.
AC 017320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Ceratitis.
OX NCBI_Taxid=7213;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96123276; PubMed=8533095;
RA Zwieler L.J., Saccone G., Zacharopoulou A., Besansky N.J.,
RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
RT "The white gene of Ceratitis capitata: a phenotypic marker for
RT germline transformation."
RL Science 270:2005-2007(1995).
CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X89933; CAAG1998.1;
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003439; ABC_transporter.
CC Pfam: PF00005; ABC_tran. 1.
CC ProDom: PD000006; ABC_transporter. 1.
CC SMART: SM00382; AAA. 1.
CC TIGRfam: TIGR00955; Jao1204. 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
DR Pfam: ATP-binding; Transmembrane; Transport.
FT NP_BIND 121 128 ATP (BY SIMILARITY).
FT TRANSMEM 427 445 POTENTIAL.

```

```

FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 507 525 POTENTIAL.
FT TRANSMEM 534 555 POTENTIAL.
FT TRANSMEM 568 586 POTENTIAL.
FT TRANSMEM 651 670 POTENTIAL.
FT CARBOHYD 628 628 POTENTIAL.
FT CARBOHYD 643 643 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 679 AA; 75145 MW; 3F9CBCT8A835C4CC CRC64;

```

Query Match 17.8%; Score 623.5; DB 1; Length 679;
 Best Local Similarity 28.3%; Pred. No. 3, 2e-39;
 Matches 169; Conservative 112; Mismatches 264; Indels 53; Gaps 8;

```

OY 88 IONSFKVRSQMLAIIGSSCGRASLDVITRGHG-CIKSGOI-WINQPSPOLVR 145
Db 104 LKNDGVAYPELLAVMSSGAGKTLTNAFAFSKQVOISPTIRLNLHPDAAKMQ 163
OY 146 KCVAVRHQNDLRLTYRETLATFAQRRLPRTSQQRQVEDVLAELRQCADTRV 205
Db 164 ARCAVQDDLEFISLAREHLFOAMVRMHRHTOKQVQRVDVQIDLSLGCQNTLI 223
OY 206 G-NMYVRGLSGERRRVSIGVQLMNPGLILDEPTSGDSFTAHNLVKYLSLAKGNRL 264
Db 224 GVGKRVKGLSGCKRKLAFASEALTDPELLCDEPTSGDSFMAHVSVOYKLKSGKKT 283
OY 265 VLISLHOPRSDIFRLFDLVILMTSGTPYICAAQHMVOYFALIGPCPRVSNPADFYVDL 324
Db 284 VILTIHQPSSLEFLFKIILMAGRAFLGTPEAVDFSYIGATCPNTYTPADFYVOV 343
OY 325 TS-----IDRSDEQL-----ATREKAOSLALFLEK--VRDLDFLWKAFTKD 367
Db 344 LAVVPGREVSRRVAKICDNFAVGKVSREMEQNFQVLVNSNGRKNDEMYTKASM-- 401
OY 368 LDEDTVESSVTPLDTNCLSPPTKMPAVQOFTLLIRQISNDFRDLPTLLIHGAECIM 427
Db 402 -----FMQFRAVLMRSMLSYLKDELVLAVKRLQTTWV 433
OY 428 SMTIGLYFGHGSIQLSFMDTAALLFMIGALIPFNVIIDVTSKCYSERAMLYELEDGLY 487
Db 434 AVLIGLIFLGOQLTQYGVNMINGAIFLFLTNMTFQNSFATITFTLPEVMEETRSLX 493
OY 488 TTPPYTPAKLGLPEHCAYIIITGMPYMWLANRPGLOPPLFLVLMVAVFCCRIMAL 547
Db 494 RCDYTFELGKIALPELPLVPEFLETAIVPELIGLRGVDFALALVTVLVANVSTSGF 553
OY 548 AAALPLTFHMASFFSNALYNSFYLAGGFMINLSLMTVPAMIKVSELR--CFEGIMAKTQ 607
Db 554 IISCASSISMALSVPPVYIIFPLFGGEFLNSGVYVYFKMLSYLSMFRYANEGILLNO 613
OY 608 FS---RRTYKMPGLNLTIAVSGDKILSAMELDYPYAIYLIYVIGSGFMYLYVSL 662
Db 614 MADVKGEITCTLSNTTQPSSEVILETINFASDLPPDFGLALLIVGFRISAYIAL 671

```

RESULT 12

```

ABCL_MOUSE
ID ABCL_MOUSE STANDARD: PRT: 666 AA.
AC 064343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
GN ABCG1 OR ABC8 OR WHRT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Son D., Arciniegas S., Wu R.;

```


OC Neoptera: Endopterygota: Diptera; Brachycera: Muscomorpha; Oestroidea:
 OC Calliphoridae: Lucilla.
 OX NCBI_TaxId=7375;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97087158; PubMed=8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.:
 RT "The structure, sequence and developmental pattern of expression of
 the white gene in the blowfly *Lucilla cuprina*.";
 RL Insect Mol. Biol. 5:251-260(1996).
 RN [2]
 RP SEQUENCE OF 490-584 FROM N.A.
 RX MEDLINE=90264941; PubMed=1971656;
 RA Ellicur A., Vacek A.T., Howells A.J.:
 RT "Cloning and characterization of the white and topaz eye color genes
 from the sheep blowfly *Lucilla cuprina*.";
 RL J. Mol. Evol. 30:347-358(1990).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U38989; AAA82057.1; -;
 DR EMBL: X53265; CA337365.1; -;
 DR InterPro: IPR003593; AAA_Arpase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran.1
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART; SM00382; AAA.1.
 DR TIGRfams; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KM Pigment: ATP-binding; Transmembrane; Transport.
 FT NP_BIND 119 126
 FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 534 554 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 647 667 POTENTIAL.
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;
 Query Match 17.7%; Score 620.5; DB 1; Length 677;
 Best Local Similarity 29.3%; Pred. NO. 5.4e-39;
 Matches 174; Conservative 119; Mismatches 259; Indels 41; Gaps 13;
 OY 88 IONLSFKRSGOMLAITSSCGRASLDVYTR-CHGCKIKSGOI-WINCOPSSPOLVR 145
 DB 102 IKWGVAVYPELLAVMGSSGAGKTTLLNALAFSARGVQISPSVRMLNGHPYAKEMQ 161
 OY 146 KCAVHRQNLPLVYRETLFAQMRLPRTSQORDRDEVIATLRLQCAQDVRV 205
 DB 162 ARCAVVOODLFIGSLTAREHLIQAQVEMPTKOKLORVDQVDDLSLKONNTI 221
 OY 206 G-NMYVRLSGGERRRISGVQLMNPGLILDEPTSGLDSTFANHLVTKTSLAKGRL 264
 DB 222 GVPGRVKLSGGERRLAFASALDPLLCIDDEPTSGLDSPMAASVQVQKLKLSQKRT 281
 OY 265 VLSIHPORSDIFRLFDVLVMTSGTPIYLCAQAHQVYPAIGTGPCRISNPADFYVDL 324
 DB 282 VILTHOPSSSELPFEDLILMAAGRAVAFGLTPEAVDVFSGFISGOCPTNPNPADFYQV 341
 OY 325 TSIDRRSREQLATREKAQSLALF-LEKV-RDLDLFLMK--AEFKDLEDEPTCESSVTP 380

DB 342 LAV---VPGREIESDRISKICDNFAYGVSRMEMQNFQIAKTDGLQKDD----- 390
 OY 381 LPTNCLPSPTKMPGAVQOFTTLIRQISDNFRLDTLLIHGAALMSTTIFLYFGHS 440
 DB 391 -ETTLIKRSMF-----TQRAIMMSWISTLEPLLVKRLIQTMVAALLGLFLNPM 445
 OY 441 IOLSEMDPAALFMGALIPFNVIIDVSKSCSEBAMLYELEDGLYTTGPFYKILGE 500
 DB 446 TQGVNMINGALFELTNTFQNVAVINVFSELPVFRKRSRLKRDYFLFKTLAE 505
 OY 501 LPEHCAYIITGMPYWLANLRPLGLPFLHFLVLVWVFCRRIMAAALPTFHMS 560
 DB 506 LPLFLVPEFLFAIYAVPMGLRPGITHEFLSALALVTLVANVSTSGYLISCSTSNAL 565
 OY 561 FESNALNYSFLAGFPMINLSLWTPVAMISVSLRCMCFEMLIQSRRTYKMPLC-- 618
 DB 566 SVGPPLTTPFLLEGGVFNLSGSPVYFKMLSTFSFRYANEGLLLNQW----DVQPGEI 621
 OY 619 -----NLTIAVSGDKILSMEID---SYPLVAYILVIGLGGFVWLYVSLR 663
 DB 622 TCTSTNTCPSSGXVXLETINRDKFTRLVGLILIL-----TFRIAGTVAAK 670
 RESULT 14
 ID ABG1_HUMAN STANDARD; PRT; 678 AA.
 AC P45844; O9BXK6; O9BXK7; O9BXK8; O9BXK9; O9BXL0; O9BXL1; O9BXL2;
 AC O9BXL3; O9BXL4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (white protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABCG8 OR WHR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Retina.
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Laliot M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the *Drosophila* white
 RL gene and mapping to chromosome 21q22.3.";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Matanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Matsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Horstschler K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesenfeld L., Dagand E.,
 RA Weinmeyer S., Borzym K., Gardiner K., Mizellic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
 RA Wattenhofer M., Gulpioni M., Barras C., Rossier C., Shibusky K.,
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness

DR EMBL; AF323645; AAK28835.1; JOINED.
 DR EMBL; AF323644; AAK28835.1; JOINED.
 DR EMBL; AF323647; AAK28835.1; JOINED.
 DR EMBL; AF323648; AAK28835.1; JOINED.
 DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 17.68; Score 617; DB 1; Length 678;
 Best Local Similarity 25.78; Pred. No. 9.8e-39;

Matches 173; Conservative 130; Mismatches 266; Indels 104; Gaps 18;

33 DNSLYFT--YSGOPN-----TLEVRDNLQVNDLASQVMPFQQLAQFMPTSPSCQSCSEL 86
 57 DNNLEAOPFSSLPRAAVNIFERDLSISV--PESPMWRKKGYKTL----- 100.
 87 GIONLSFKVRSQOMLAIIGSSGGRASLDVITGRGHGKIKSGQIWMINGOPSSPOLYVR 146
 101 -LKGISGRNSELVAIMPSCGKSTLNNIAGYETG--KGNVLLINGLRPDLRCFKR 157
 147 CVAHVRQHNQLLPNTLVRETLAFIAOMRLPRTFSQAQRDKRVEDVIAELRLQCADTRVG 206
 158 VSCYIMODMLPHTLVQEAAMVSAHLKIQE--KDEGRREMYKEILLTALGLLSCANTRTG 215
 207 NMYVRLSGGERRRVSIGVOLLMPGILLDEPTSGDSTFAHNLVKTLSLAKGNRLV 266
 216 S-----LSGGQRRKRLAIETLVNPPVMEFDEPTSGDASCFQVYSLKKGAGGGRSII 270
 267 ISLHOPRSDIFRLFDVLLMTSGTPYILGAQHMVQYFAIGPCPRYSNPADFYVDLNS 326
 271 CIHQPSAKLFELFQLYLVLSGGQCYRKNVCNLPVLLDGLNCTYINPADFYVEVAS 330
 321 IDRSRDEQL--ATEKQASLALFLEKVRDI-----DFTMKAEI-----KDL 369
 331 GEGYDONSRLVAVRE-----GMCDSDHRRDLGDAEVPFLMHRPSEVQKRLKGLR 385
 370 EDTCYESSVTPDTCCLPSPKMGAVNOOFTLLIRROISNDPRDLPTLLIHGAECIMSM 429
 386 KDSSMEGCHFSASCL-----TQFCILFKRTPLSIMDSVHLRLRTSHIGIL 435
 430 TIGFLYFPHGSIOLSFMDTALLF-----MIGALIPFNVIIDVISKYSEBAMLYELE 483
 436 LIGLILYIGINPAKKVLSNSGFLPFSMLFLMFAALP-----TVLFLPE 480
 464 DGL-----YITGPFPAKILGELPEHCAYIIITGMPYTWLANLRGLOPFLIHFL 534
 481 MGVFLREHLNWTYSLKAYVLAKTADVDPQIEMFPAVCSIVYMTSOPSADAVFVLEAL 540
 535 VWLVFFCCRIALAAALLPFFHMASFFSNALYNFYLAGFMNIMLSLMTVPAMISKVS 594
 541 GTMISLVAOSLGLLIGASTSLQVATFVGPTATFVLLFSGFVSFDTPTVLLQWMSYIS 600
 595 FLRMCFEGIMKIQF--SRRTYKMPILNLTIAVSGDKILSAMELDYSPLVAILIYIGLSG 652
 601 YVRVGFEGEVIILSIYGLDREDLHCDIDETCHPOKSAIILRELDVENAKLY-LDFIYLG-- 656
 653 GPMVLYYSLRPI 665
 657 ----LFTLSRLTI 665

RESULT 15
 YPC3_CAEEL
 ID YPC3_CAEEL STANDARD: PRT: 598 AA.
 AC Q11180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative ABC transporter C05D10.3 in chromosome III.
 GN C05D10.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Du Z.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC EMBL; U13645; AAA20989.2;
 DR Wormpep: C05D10.3; CR29170.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRfam: TIGR00955; 3a01204; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; FALSE NEG.
 DR PROSITE: PS00933; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
 FT TRANSMEM 27 34
 FT NP_BIND 336 356
 FT TRANSMEM 425 445
 FT TRANSMEM 453 473
 FT TRANSMEM 478 498
 FT POTENTIAL.
 FT POTENTIAL.
 SQ SEQUENCE 598 AA: 66906 MW: 9D6414E06898E343 CRC64;

Query Match 17.18; Score 600; DB 1; Length 598;
 Best Local Similarity 27.98; Pred. No. 1.6e-37;
 Matches 170; Conservative 116; Mismatches 260; Indels 64; Gaps 15;

88 IONLSFKVRSQOMLAIIGSSGGRASLDVITGRGHGKIKSGQIWMINGOPSSPOLYVR 147
 10 LHNVSMAESGKLLAIIGSSGGRKTTMNVLSRLTNDVGSSTLIDRRANKKIREM 69
 148 VAHVRQHNQLLPNTLVRETLAFIAOMRL-PTFSQAQRDKRVEDVIAELRLQCADTRVG 206
 170 SAFVOOHMFVGTMTAREHLQFMARLRMGDQYYSHERQLRVEQVLTQWGLKKCADTVIG 129
 207 -NMYVRLSGGERRRVSIGVOLLMPGILLDEPTSGDSTFAHNLVKTLSLAKGNRLV 265
 130 IPNQLKSGSGEKKRLSPASELTLPKILFCDEPTSGDAPAGVVALRSLAONGMTV 189
 266 LISHQPRSDIFRLFDVLLMTSGTPYILGAQHMVQYFAIGPCPRYSNPADFYVDLT 325
 190 ITIHQPSHHVYSLFNVCMAACGRVITLGRDQAVPPEKCGYCPRAYNPADHLIRL 249
 326 SIDRSRDEQLATREKKAOSLALFLEKVRDLDPLMKAEYTDLDTC-----VES 376
 250 AVIDSDDRAVTSMT-----ISKIR--QGFV-----STDLGQSVLAIGNANKLRAAS 292
 377 SVTPDTCCLPSPKMG-----PGAVOQFTLLIRROISNDPRDLPTLLIHGAECIMSM 429
 293 FVTGSDTS-----EKTTFPNODYNSFWTQFALFERSMLTVIRDPNLLSVRLQTLITAF 349
 430 TIGFLYFPHGSIOLSFMDTALLFMIKALIPFNVIIDVISKYSEBAMLYELE 481
 350 ITGIVFF-----QPTVPTATISINSIM-FNHRNNKMKLQPPNVVITAEPIYARE 401
 482 LEDGLITGPFPAKILGELPEHCAYIIITGMPYTWLANLRGLOPFLIHFLVLYVFC 541

Db 402 NANGYRTSAIFLAKNIAPOLYIILPTILYNTIYVMMGSLYPNEMNYCFASLVTLLITNV 461
OY 542 CRIMALAAALLPFFHMASFFSNALYNSFLAGFEMINLSSLMTPWPAISKVSPLRWCFE 601
Db 462 AISISYAVATIFANTDVAMTILPIFVVPIMAFGGFEFTEDAIPTSEKWLSSLSYFKGYE 521
OY 602 GLM-----KIQFSRRTYKMPGLNLT---AVSGDKILSAMELD-SYPLXAYLYIVIGLSG 652
Db 522 ALAINEMDSIKVIEPCFNSSWTAFALDSCPKNGHQVLESIDFSASHKIEDI-SILPGMFI 580
OY 653 GFNVLYYVSL 662
Db 581 GIRIIRAYVAL 590

Search completed: July 25, 2003, 17:12:45
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein. - protein search, using sw model

Run on: July 25, 2003, 16:49:39 ; Search time 86 Seconds

(without alignments)
1242.128 Million cell updates/sec

Title: US-09-989-981a-8

Sequence: 3506 1 MACKAAEENGLPKGATPQDT.....FVLYVSLFIRKQPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3506	100.0	673	24	AAE31705 Human ABCG8 protei
2	3502	99.9	673	23	ABP52129 Homo sapiens ABC t
3	2888.5	82.4	672	24	AAE31703 Mouse ABCG8 protei
4	1961	55.9	374	23	ABG61539 Human transporter
5	730.5	20.8	632	21	AAE18079 Arabidopsis thailia
6	730.5	20.8	648	21	AAE18078 Arabidopsis thailia
7	724	20.7	625	21	AAE18080 Arabidopsis thailia
8	713	20.3	652	23	AAU96986 Rat ABCG5 protein.
9	705	20.1	651	23	AAU96990 Human ABCG5 mutant

10	697	19.9	651	23	AAU96984 Human ABCG5 protei
11	697	19.9	651	23	AAU96993 Human ABCG5 mutant
12	697	19.9	651	23	AAE13290 Human siltosterolae
13	697	19.9	651	24	AAE31704 Human ABCG5 protei
14	696	19.9	651	23	AAU96989 Human ABCG5 protei
15	694	19.8	651	23	AAU96992 Human ABCG5 protei
16	691.5	19.7	652	23	AAU96985 Human ABCG5 protei
17	688.5	19.6	652	23	AAE13289 Mouse siltosterolae
18	688.5	19.6	652	23	AAE13308 Mouse siltosterolae
19	688.5	19.6	652	23	AAE13309 Mouse siltosterolae
20	688.5	19.6	652	24	AAE31702 Mouse ABCG5 protei
21	675	19.3	649	23	ABP52128 Homo sapiens ABC t
22	666	19.0	657	23	ABP52127 Murine BCRP (mBCRP
23	665	19.0	667	21	AAE18961 Drosophila melanog
24	656	18.7	667	22	ABP59384 Human transporter pr
25	642.5	18.3	655	22	AAE60104 Human BCRP. Homo
26	642.5	18.3	655	23	AAU80028 Human BCRP. Homo
27	642.5	18.3	655	23	AAU14781 Human BCRP. Homo
28	640.5	18.3	655	23	AAU80029 Human ABCG2 protei
29	640.5	18.3	653	20	AAE15221 Human ABCG2 protei
30	638.5	18.2	655	21	AAU95365 Breast Cancer Resl
31	638.5	18.2	655	22	AAU04348 ATP-binding cassel
32	638.5	18.2	655	23	ABP52127 Homo sapiens ABC t
33	638.5	18.2	655	23	ABP52127 Homo sapiens ABC t
34	638.5	18.2	655	23	AAU14782 Human BCRP (hBCRP
35	638.5	18.2	655	23	AAU14783 Human BCRP (hBCRP
36	634.5	18.1	666	23	ABP57112 Mouse ischaemic co
37	621	17.7	666	23	ABP57112 Human ABC transpor
38	620	17.7	666	23	ABP57112 Human ABC transpor
39	618	17.6	638	23	ABP58349 Human ABC transpor
40	617	17.6	674	23	ABP52126 Homo sapiens ABC t
41	612	17.5	408	23	AAU96991 Human ABCG5 protei
42	610.5	17.4	609	22	ABP64566 Drosophila melanog
43	572.5	16.3	646	24	ABP96994 Amino acid sequenc
44	571.5	16.3	646	23	AAE28968 Human ABCG4 transp
45	570.5	16.3	646	24	ABP96996 Amino acid sequenc

ALIGNMENTS

RESULT 1

AAE31705 standard; Protein; 673 AA.

XX	AC	AAE31705;	
XX	DT	24-MAR-2003 (first entry)	
XX	DE	Human ABCG8 protei.	
XX	KW	ABC family cholesterol transporter; ABCG8; sterol-related disorder;	
XX	KW	siltosterolemia; hyperlipidaemia; hypercholesterolemia; gall stone;	
XX	KW	HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;	
XX	KW	human; ATP-binding cassette; siltosterolemia susceptibility gene; SSG;	
XX	OS	Homo sapiens.	
XX	PN	WO200281691-A2.	
XX	XX	17-OCT-2002.	
XX	PF	20-NOV-2001; 2001WO-US43823.	
XX	PR	20-NOV-2000; 2000US-252235P.	
XX	PR	28-NOV-2000; 2000US-253645P.	
XX	PA	(TUL) TULARIX INC.	
XX	PA	(TEXA) UNIV TEXAS SYSTEM.	
XX	PI	Hobbs HH, Shan B, Barnes R, Tian H;	

DR WPI; 2003-058548/05.
 DR N-PSDB; AAD48883.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating
 PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies -
 XX
 PS Claim 22; Page 81-82; 94pp; English.

XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is human
 CC ABCG8 protein.
 CC
 XX

SQ Sequence 673 AA:

Query Match 100.0%; Score 3506; DB 24; Length 673;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 DB 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 QY 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 DB 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 QY 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFIAQMLPRPFS 180
 DB 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFIAQMLPRPFS 180
 QY 181 QAOQRKVEDYIAELRLQCADTRGVNMYRGISGGERRRVSGVOLLMPGILIDDEPT 240
 DB 181 QAOQRKVEDYIAELRLQCADTRGVNMYRGISGGERRRVSGVOLLMPGILIDDEPT 240
 QY 241 SGIDSTFAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLMTSGTPYLCAAOHM 300
 DB 241 SGIDSTFAHNLVKTLSRLAKGNRLVLSIHOPRSIDIFRLFDVLMTSGTPYLCAAOHM 300
 QY 301 VOYFAIGYPCPRYSNPADFYDLTSDRSREOELATREKAQSLAALFEKVRDLDEL 360
 DB 301 VOYFAIGYPCPRYSNPADFYDLTSDRSREOELATREKAQSLAALFEKVRDLDEL 360
 QY 361 WKAETKRDDEDCVSSWTPDLTNCPLPSTKMPGAVOFTLIRROISNDFRDLPTLLH 420
 DB 361 WKAETKRDDEDCVSSWTPDLTNCPLPSTKMPGAVOFTLIRROISNDFRDLPTLLH 420
 QY 421 GAELACIMSWGTFELFGHGSIQLSFMDTAAALFMGALIPFNVLIDVTSKCYSEKAMLY 480
 DB 421 GAELACIMSWGTFELFGHGSIQLSFMDTAAALFMGALIPFNVLIDVTSKCYSEKAMLY 480
 QY 481 ELEDGLYTTGPFYFAKILGELPERCAVYIIIGMPTYWLANRPGLOPELHFLVWLVF 540
 DB 481 ELEDGLYTTGPFYFAKILGELPERCAVYIIIGMPTYWLANRPGLOPELHFLVWLVF 540
 QY 541 CCRIMAAALAAALLPFNHASFPSNALYNSFYLAGFMALNSLWTPVPMISKVSFLRMCF 600
 DB 541 CCRIMAAALAAALLPFNHASFPSNALYNSFYLAGFMALNSLWTPVPMISKVSFLRMCF 600
 QY 601 EGLMKIQFSRRTYKAPLGNLTIAVSGDKILSAMELDSTPLVAILIYIGLSGGFVWLYV 660
 DB 601 EGLMKIQFSRRTYKAPLGNLTIAVSGDKILSAMELDSTPLVAILIYIGLSGGFVWLYV 660
 QY 661 SLRFTKQKPSQDW 673
 DB 661 SLRFTKQKPSQDW 673

RESULT 2
 ID ABP52129 standard; Protein; 673 AA.
 AC ABP52129;

10-OCT-2002 (first entry)

Homo sapiens ABC transporter ABCG8 protein SEQ ID NO:81.

ATP-binding cassette transporter; ABC transporter; modulation; D loop;
 cancer; bacterial infection; fungal infection; protozoal infection;
 antibacterial; fungicide; protozoacide.

Homo sapiens.

EP1217066-A1.

26-JUN-2002.

21-DEC-2000; 2000EP-0870316.

21-DEC-2000; 2000EP-0870316.

(UYGE-) UNIV GENT.

WPI; 2002-550404/59.

Modulating activity of ATP-binding cassette (ABC) transporters by
 influencing dimerization of nucleotide binding domains through use of D
 loop sequence of an ABC transporter, or its antisense peptide or
 peptide mimetic -
 Disclosure; Fig 3; 29pp; English.

The present invention describes a method (M1) for modulating the activity
 of ABC-binding cassette (ABC) transporters by influencing the
 dimerization of the nucleotide binding domains comprising using: (a) a
 polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop
 sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP
 consisting of the D loop sequence of (a) or (b); (c) a peptide
 mimetic or antisense peptide of (a) or (b); ABC transporters have
 antibacterial, fungicide and protozoacide activities. (M1) is useful for
 selectively modulating the activity of ABC transporters belonging to the
 group of multidrug transporter/P-glycoproteins. Bacterial, fungal or
 protozoal ABC transporters are involved in the infection of a mammal or
 in the induction of resistance to antibiotics or drugs in a mammal. (M1)
 is useful for preventing, treating or alleviating diseases associated
 with functionality of an ABC transporter. ABP52092 to ABP52140 represent
 ABC transporter proteins given in the exemplification of the present
 invention.

SQ Sequence 673 AA:

Query Match 99.9%; Score 3502; DB 23; Length 673;

Best Local Similarity 99.9%; Pred. No. 0;
 Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 DB 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 QY 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 DB 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 QY 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFIAQMLPRPFS 180
 DB 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFIAQMLPRPFS 180
 QY 181 QAOQRKVEDYIAELRLQCADTRGVNMYRGISGGERRRVSGVOLLMPGILIDDEPT 240

```

DB      |||||||
181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIGVOLLNPGILILDEPT 240
OY      241  SGLDSEFTAHNLYKTSLRLAKGNRLVLSLHQRSDIFRLFDVLLMTSGPTIYLGAOQM 300
DB      241  SGLDSEFTAHNLYKTSLRLAKGNRLVLSLHQRSDIFRLFDVLLMTSGPTIYLGAOQM 300
OY      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRSREOELATREKAOSLAALFEKVRDLDL 360
DB      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRSREOELATREKAOSLAALFEKVRDLDL 360
OY      361  WKAETKRDDEDTCEVSSVTPDNTCLPSPTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
DB      361  WKAETKRDDEDTCEVSSVTPDNTCLPSPTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
OY      421  GAELCLMSWTIGFLYFGHSIQLSFMDTAAALFMICALIPFNVLIDVISKCYSERAMLY 480
DB      421  GAELCLMSWTIGFLYFGHSIQLSFMDTAAALFMICALIPFNVLIDVISKCYSERAMLY 480
OY      481  ELEDGLYTTGPFYFAKILGELPEHCAYIIITGMPYWLNLNRPGLDPLHLFLVWLV 540
DB      481  ELEDGLYTTGPFYFAKILGELPEHCAYIIITGMPYWLNLNRPGLDPLHLFLVWLV 540
OY      541  CCRIMALAAALLPFFHNASFESNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
DB      541  CCRIMALAAALLPFFHNASFESNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
OY      601  EGLMKTQPSRRTYKMPGLNLTAVSGDKILSMELDSTPLVAILYIVIGLSGGMVLYV 660
DB      601  EGLMKTQPSRRTYKMPGLNLTAVSGDKILSMELDSTPLVAILYIVIGLSGGMVLYV 660
OY      661  SLRFKOKPSQDM 673
DB      661  SLRFKOKPSQDM 673

RESULT 3
AAE31703
ID  AAE31703 standard; Protein: 672 AA.
AC  AAE31703;
XX
XX  24-MAR-2003 (first entry)
DE  Mouse ABCG8 protein.
XX
XX  ABC family cholesterol transporter; ABCG8; sterol-related disorder;
XX  sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
XX  HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
XX  mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
XX  ABCG5.
OS  Mus sp.
XX
XX  Key Location/Qualifiers
XX  Misc-difference 440 /note- "Encoded by AAG"
XX  WO200281691-A2.
XX  17-OCT-2002.
XX
XX  20-NOV-2001; 2001WO-US43823.
XX
XX  20-NOV-2000; 2000US-252235P.
XX  28-NOV-2000; 2000US-253645P.
XX
XX  (TULAR) TULARIK INC.
XX  (TEKA) UNIV TEXAS SYSTEM.
XX
XX  Hobbs HH, Shan B, Barnes R, Tian H;
XX  WPI; 2003-058548/05.

```

```

DR      N-PSDB; AAO48881.
XX
XX  New ABCG8 polypeptides and nucleic acids, useful for treating
PT  sterol-related disorders e.g. sitosterolaemia, hypercholesterolaemia,
PT  hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT  nutritional deficiencies -
XX
XX  Claim 22; Page 76; 94pp; English.
CC
CC  The invention relates to ATP-binding cassette (ABC) family cholesterol
CC  transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC  provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC  known as sitosterolaemia susceptibility gene (SSG). Sequences of the
CC  invention are useful for treating or preventing sterol-related disorders
CC  such as sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall
CC  stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC  They are also useful in gene therapy. The present sequence is mouse
CC  ABCG8 protein.
XX
XX  Sequence 672 AA:
SQ
Query Match 82.4%; Score 2888.5; DB 24; Length 672;
Best Local Similarity 81.9%; Pred. No. 3.7e-287;
Matches 551; Conservative 53; Mismatches 68; Indels 1; Gaps 1;
OY      1  MAGKAERGLPKGATPDTSGLQDRLESSESNSLYFTYSGOPTLEVRDINYOVDLAS 60
DB      1  MAEKTEETQNLNNGVYLDASGLDLSFSESNSLYFTYSGQSNLTLEVRLTYQVDIAS 60
OY      61  QVWFEBQLAQFEMPTSPSCONSCELGTONLSFKYRSGMLAIGSSGGRSLDYVYG 120
DB      61  QVWFEBQLAQFEMPTSPSCONSCELGTONLSFKYRSGMLAIGSSGGRSLDYVYG 120
OY      121  RGHGGRKSGQIMINGPSSPOLYKRCVAVHQHQLNLTIVRETLAFIAQMRLPRTFS 180
DB      121  RGHGGRKSGQIMINGPSSPOLYKRCVAVHQHQLNLTIVRETLAFIAQMRLPRTFS 180
OY      181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIGVOLLNPGILILDEPT 240
DB      181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIGVOLLNPGILILDEPT 240
OY      241  SGLDSEFTAHNLYKTSLRLAKGNRLVLSLHQRSDIFRLFDVLLMTSGPTIYLGAOQM 300
DB      241  SGLDSEFTAHNLYKTSLRLAKGNRLVLSLHQRSDIFRLFDVLLMTSGPTIYLGAOQM 300
OY      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRSREOELATREKAOSLAALFEKVRDLDL 360
DB      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRSREOELATREKAOSLAALFEKVRDLDL 360
OY      361  WKAETKRDDEDTCEVSSVTPDNTCLPSPTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
DB      361  WKAETKRDDEDTCEVSSVTPDNTCLPSPTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
OY      421  GAELCLMSWTIGFLYFGHSIQLSFMDTAAALFMICALIPFNVLIDVISKCYSERAMLY 480
DB      421  GAELCLMSWTIGFLYFGHSIQLSFMDTAAALFMICALIPFNVLIDVISKCYSERAMLY 480
OY      481  ELEDGLYTTGPFYFAKILGELPEHCAYIIITGMPYWLNLNRPGLDPLHLFLVWLV 540
DB      481  ELEDGLYTTGPFYFAKILGELPEHCAYIIITGMPYWLNLNRPGLDPLHLFLVWLV 540
OY      541  CCRIMALAAALLPFFHNASFESNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
DB      541  CCRIMALAAALLPFFHNASFESNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
OY      601  EGLMKTQPSRRTYKMPGLNLTAVSGDKILSMELDSTPLVAILYIVIGLSGGMVLYV 660
DB      601  EGLMKTQPSRRTYKMPGLNLTAVSGDKILSMELDSTPLVAILYIVIGLSGGMVLYV 660
OY      661  SLRFKOKPSQDM 673
DB      661  SLRFKOKPSQDM 673

```

RESULT 4
ABG61539
ID ABG61539 standard; Protein: 374 AA.

AC ABG61539;

XX 27-AUG-2002 (first entry)

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CDL.

XX Human; transporter and ion channel; TRICH; transport disorder;
XX neurological disorder; muscle disorder; immunological disorder; cancer;
KW scleroderma; systemic lupus erythematosus; allergy; leukaemia;
KW cell proliferative disorder; cervical cancer; breast cancer;
KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
KW myotonic dystrophy; catatonias; endocrine disorder; diabetes;
KW Grave's disease; gastrointestinal disorder; Crohn's disease;
KW renal disorder; Good pasture's syndrome; viral infection; cirrhosis;
KW bacterial infection; fungal infection; parasitic infection;
KW protozoal infection; helminthic infection; cardiovascular disorder;
KW atherosclerosis; hepatic disease.

XX Homo sapiens.

PN W0200240541-A2.

XX 23-MAY-2002.

XX 25-OCT-2001; 2001WO-US46055.

XX 27-OCT-2000; 2000US-243989P.

XX 03-NOV-2000; 2000US-245904P.

XX 09-NOV-2000; 2000US-247673P.

XX 17-NOV-2000; 2000US-249661P.

XX 20-NOV-2000; 2000US-252232P.

XX 01-DEC-2000; 2000US-250790P.

XX (INCYTE GENOMICS INC.

XX Tang YT, Yue H, Nguyen DB, Hafalia AJA, Elliott VS, Lu Y;

XX Walla NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjapala M;

XX Rankumar J, Aryizu C, Gietzen KJ, Lal PG, Azimzal Y, Khan FA;

XX Thangavelu K, Thornton M, Lu DM, Tribouley CM, Warren BA;

XX Ison CH, Das D, Kaumann BE, Policky JL, Kearney L;

XX WPI; 2002-463570/49.

XX N-PSDB; ABK83218.

XX New transporters and ion channels (TRICH) polypeptides, useful for

XX diagnosing, preventing, and treating disorders associated with an

XX abnormal expression or activity of TRICH, e.g. immunological, muscular

XX or renal disorders

XX Claim 1; Page 143-144; 178pp; English.

XX The invention relates to human transporters and ion channels (TRICH)
XX polypeptides, a naturally occurring amino acid sequence 90 & identical to
XX TRICH, a biologically active fragment of TRICH or an immunogenic fragment
XX of TRICH. Also included are an isolated polynucleotide encoding TRICH,
XX a recombinant polynucleotide comprising a promoter sequence operably
XX linked to the TRICH polynucleotide, a cell transformed with the
XX recombinant polynucleotide, a transgenic organism comprising the
XX recombinant polynucleotide, an isolated antibody that binds specifically
XX to TRICH, and screening for compounds which bind to TRICH, modulate
XX TRICH, modulate TRICH expression or are ant/agonists of TRICH.
XX The polypeptides are useful for diagnosing, treating, and
XX preventing transport, neurological, muscle, immunological disorders
XX (e.g. scleroderma, systemic lupus erythematosus, allergies), cell
XX proliferative disorders such as cancers (e.g. leukaemia, cervical or
XX breast cancers), neurodegenerative disorders (e.g. Parkinson's disease,
XX Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy,
XX catatonias), endocrine disorders (e.g. diabetes, Grave's disease),

CC gastrointestinal disorders (e.g. Crohn's disease), renal disorders
CC (e.g. Good pasture's syndrome), viral, bacterial, fungal, parasitic,
CC protozoal and helminthic infections, cardiovascular disorders (e.g.
CC atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many
CC other diseases and disorders detailed in the specification. They can also
CC be used in assessing the effects of exogenous compounds on the
CC expression of nucleic acid and amino acid sequences of transporters and
CC ion channels. TRICH or its fragments may also be used in screening for
CC compounds that specifically bind to and modulate the activity of TRICH.
CC The polynucleotides can be used to create knock-in humanised animals or
CC transgenic animals to model human disease. The present sequence
XX represents a TRICH protein.

XX Sequence 374 AA:

Query Match 55.9%; Score 1961; DB 23; Length 374;

Best Local Similarity 99.7%; Pred No. 3e-192; Indels 0; Gaps 0;

Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 300 MWYFTTAIGYPCPRYSNPADFYVDLTSIDRSRQDELATREKASIALAFLEKVDLDF 359

Db 1 MWYFTTAIGYPCPRYSNPADFYVDLTSIDRSRQDELATREKASIALAFLEKVDLDF 60

QY 360 LMKATKDLDEDTCESSVTPDTRNCLSPTRKMPGAVQFTTLIRQISNDFRDLPTLLI 419

Db 61 LMKATKDLDEDTCESSVTPDTRNCLSPTRKMPGAVQFTTLIRQISNDFRDLPTLLI 120

QY 420 HGAECALMSMTIGFPGHSGISLSPMTALLPFGALIPENYLDVYSQCSRAALY 479

Db 121 HGAECALMSMTIGFPGHSGISLSPMTALLPFGALIPENYLDVYSQCSRAALY 180

QY 480 YELEDGLYTTGYPYFAKILGELPERCAVYIIIGMPTVLANLRGLOPFLHFLVWLVY 539

Db 181 YELEDGLYTTGYPYFAKILGELPERCAVYIIIGMPTVLANLRGLOPFLHFLVWLVY 240

QY 540 FCCRIMAAALALPTFMASFNSALNYSYTLAGEFINUSLMTVPAMISKVSFLRWC 599

Db 241 FCCRIMAAALALPTFMASFNSALNYSYTLAGEFINUSLMTVPAMISKVSFLRWC 300

QY 600 FEGLMKIOFSRRTKMPGNLTIVASGDKILSAMELDSYPLATYLYIGSGFMYLYX 659

Db 301 FEGLMKIOFSRRTKMPGNLTIVASGDKILSAMELDSYPLATYLYIGSGFMYLYX 360

QY 660 VSLRFIRKQPSQDW 673

Db 361 VSLRFIRKQPSQDW 374

Db 361 VSLRFIRKQPSQDW 374

RESULT 5

AAAG18079

AAAG18079 standard; Protein: 632 AA.

AAAG18079;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 19344.

Protein identification: signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;

termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

PR 22-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140333.
PR 24-JUN-1999; 99US-0140334.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.8%; Score 730.5; DB 21; Length 632;
Best Local Similarity 30.7%; Pred. No. 2e-65;

Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

OY 9 RGLKRGARPOPTSGDRLSESES--DNSLYFTTSGCPNTLEVDLNTQVDLACQVWFE 66
DB 3 QGLDMSDTQSKSVLAFFPTTSGLOMSY-----PITKFEVYVKVI-----E 49
OY 67 QLAOFKMPWTPSPSCNCELGIONLSFKVRSGOMLAISSGCRASLLDVTGRGHGX 126
DB 50 QTSOCMSMKSE-----KTLNGITGVCPCGFELAMGPGSGKTTLSLGR--LSK 102
OY 127 IKSGQIWINQOPSSPOLVRKCAVAVRHQNLPLVRETLAFLAOMLPTTFSSOARDK 186
DB 103 TFSCKVMYNGQPFSGCIKRR-TGFVADDDVLYPHLFWETLEFALLRLPSLIRDEKAE 161
OY 187 RVEDVIAELRLROCADRVGMVYRGLSGERRRVSIGVOLLMPGIIIDEPISGDSF 246
DB 162 HDVAVIAELGLMKTNSMIGPLRGISGGEKKRVSIGOEMLIMPSTLLIDEPISGDSF 221
OY 247 TANHVLVTLRLAGNRLVLSLHQPSPDIFRLFDVLMTSGPTVIGAAQHWVQFETA 306
DB 222 TANHIVTTIKRLAGGRVYVTTIHQPSRIYHMDKVVLSGSPYIYGAASSAVERFSS 281
OY 307 IGYCCPRYSNPADFYVDLTS-----IDRSREDELATREKASIALAFLKVRDLDFLW 361
DB 282 LGFSTSLVNPADLLDLANGIIPDQKETSEQKTVK--ETLYSAVEKNI----- 331
OY 362 KAEKRLDEDCVSS---VTPIDTNCUSPTMPGAVOQFTLIRQI--SNDFRLDPT 416
DB 332 --STK-LKAEICNAMESHYETKAAKNLSEOMCTWMOFTVLLRGVAREERFESFNK 388
OY 417 LLIHGAENCLMSMTIGELYFGHSIQLSFMDTALLFMIGALLPENVILDVISCYSERA 476
DB 389 LRIF---QVISVALGGLMWH--TPKSHIDRTALLFFSVWGFYLYNAVFPPOEKR 444
OY 477 MLYTELEDGLYTTGPFYFAKILGELPERCAVYIIYGMPTVLANLRPGLPFTLHFLYV 536
DB 445 MLIKERSSGMYRLSYFARMVGDLPLELAPFAVFIYIYMGGLKDPFTFILLSLVYL 504
OY 537 LVVCCRRMLAALALPFTFHMASFSPNALXNSYLAGGMINLSLMTVP---AMISKV 593

DB 505 YSLVAOGLAFLGALLNNIKOATTLASVTLIVFLIAGGYVQ-----QIPPTVWLYL 559
OY 594 SELKCFEGLMKIOPSRTTY-----KMPLGNTTAVSGDKTILSANEL 635
DB 560 SYSTYCYKLLGIOTDDDIYECSCGVKCRGDDPFAIKSMGLNMDI---DVEFMGYML 615
OY 636 DSYPLAIVLYIGISGGFVNLVYVSLR 663
DB 616 VGYRLMA-----YMLHRYKLR 632

RESULT 6

ID AAG18078 standard; Protein; 648 AA.

AC AAG18078;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19343.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143847.
PR 13-JUL-1999; 99US-0143847.
PR 14-JUL-1999; 99US-0143824.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145222.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.8%; Score 730.5; DB 21; Length 648;
Best Local Similarity 30.7%; Pred. No. 2,1e-65;
Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145108.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145216.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148177.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151067.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159337.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 20.7%; Score 724; DB 21; Length 625;
 Best Local Similarity 30.9%; Pred. No. 9, 3e-65;
 Matches 211; Conservative 113; Mismatches 258; Indels 100; Gaps 20;

15 ATPQDTS--GADRLFSSESNSLYFTYSCGPNTEVDLNVQVDLASQVWFEOIAQK 72
 11 AFRITTSQPLQ-----MSMY-----PITLKEFEVYKVKI-----EDTSCM 48
 73 MPWTSPSCQNCCEJONLSTFKVNSGOMALITSSGCGRASLDVITGRGHGKIKSGOI 132
 49 GSKMSKE-----KTLNGLITGMVCPGEFLAMGSGSKTLLSMIGR--LSMTFSKV 101
 133 WINGOPSSPOLVRVCANVHROHNDLPLVTRTELAFIAQNRLEPTFSQADRKREVEVI 192
 102 MYNQPFSSGCIKRR--TGFVAODDVLYPHLTWETLFFYALLRLPSSILTRDEKAEHVPI 160
 193 AELRLQCADRVGNMVRGLSGGERRRVSIGVOLLNPGILIDEPSTGDSFTAHNV 252
 161 AELGLNCTNMIGGPLERGISGGEKRVISIGCEMLNPSILLDEPTSGDSTTAHRIV 220
 253 KTLRLAKNRVLVLSIHOPSRDIPLRFDVLVLMTSGTPIYLGAQHNVQYFTAIQVPCP 312
 221 TTIRKLASGGRTVYTTTHIOPSSRIYHMDKVLLSEGPSIYYGAASSAVEYFSSIGSTS 280
 313 RYSNPADFYVDLTS-----IDRRSREQELATREKAQSIALFLKRVRLDDFLKAEKTD 367
 281 LTVNPADLLDLANGIPDPDOKETSEDOEKVK--ETLVSAVEKENI-----STK- 327
 368 LDEDTCVES--VPLDTNCLPSPTKMPGAVOOFLLHROI--SUDFRLDPTLLHGA 422
 328 LKAEICNAESHSTYTTAAKNLKSSEOMCTTWTQFVLLQCKVREKRFESFNKLRIP-- 385
 423 EACLSMTIGLFLFGHCSIQLSFMDTAALEMGALIPFNVLIDVISCYSEKRAALYEL 482
 386 -QVTSVAFILGLIMWH--TPKSHIODRTALLFEFFSVGFPLVNAVFTFPOEKMLKER 443
 483 EDGLYTTGPEYFAKILDELBHCYIITIGMPYTWLANLRBGLOPFLHLFLVLVVECC 542
 444 SSGMYRLSSYFMANNVDPLELALPTAFVFIITMGMGLKRPDDPTFTLSLVLYSVIVA 503

QY 543 RINALAAALLPTFHMAFFSNALYNSEYLAGFMILSLMTWTP---AMISRVSLRMC 599
 Db 504 QGGLGALFALMLMIKQATITLAVTTLVFLIAGGYVO-----QIPPIVWLKILSYCYC 558
 QY 600-FEGIMKIQESRRY-----KMPICNLTIAVSGDKILSAMELDXYPLX 641
 Db 559 YKLLIGIYTDYDDYECSSKVCWCRVGFPAIKSMGLNNLMI-----DYFVAGVNLVGRIM 614
 QY 642 AIIYIYIGLSGGFMVLYYSLR 663
 Db 615 A-----YMALHRYKLR 625
 RESULT 8
 AA096986
 ID AA096986 standard; Protein; 652 AA.
 AC AA096986;
 XX 30-JUL-2002 (first entry)
 DE Rat ABCG5 protein.
 KM Rat; ABCG5: ATP-binding cassette gene 5: sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 OS Rat sp.
 XX MO200227016-A2.
 PN 04-APR-2002.
 PD 25-SEP-2001; 2001MO-US29859.
 PF 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 PI WPI: 2002-416483/44.
 DR N-RSDB; ABRK1686.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX Example 3; Page 45; 66pp; English.
 PS The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention.
 SO Sequence 652 AA:

Query Match 20.3%; Score 713; DB 23; Length 652;
 Best Local Similarity 30.0%; Pred. No. 1,3e-63;
 Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;
 QY 12 PKGAT-PQDSGIGQDRLFFSESSNSLFTYSGGPNLLEFVRLDYQVDIASQV-PREPOLA 69
 Db 9 PEGARGHNRRSQ-----SSLEGGV--TGESEARHSIGV--LNVSPSVSNRVPW----- 55
 QY 70 QFKMPTSPSCQNSCEIGI-ONLSFKVRSQGMALIGSSGCGASLIDVTYTGSGHGKIK 128
 Db 56 -----WNKSCQQRKMDKILKADVSLYIESGQIMCIISSSGSKTLLDASGHLRTGTL 110
 QY 129 SGQIMWINGOPSSPOLYRKCAVAVRQHNQLLPNLTVETLAFIAQMLRPTFSQAQRDKRY 188
 Db 111 EGEVFNCGELRBDQPODCVSYLLQSDVFLSLSLVETRLRYTMALV-RSSADYFDKIV 169
 QY 189 EDVIAELRLQCADTRYQNMKYVGLSGGERRRVSIGVQLMNPGLIILEPPSGLSFTA 248
 Db 170 EAVLTETLSHVADOMIGNVNFQGISGERRRVSIAAQLQDPRKVMVLDPEPTGLDCMTR 229
 QY 249 HNLVKTLSRLAKGNRLVILSLHOPRSDIFRLPDVILMTSGPTIYIGAAQHWQVPTATG 308
 Db 230 NHTVLLVLELARRNRIVYTIHOPRSELEFHFQKIALITYGLVFCGTPPEMLGFENNCG 289
 QY 309 YPCPRYSNPADFYVDLTSIDRSREQLATREKAQSLAALF-----LEKVRDL 356
 Db 290 YPCPRHSNPDYFMDLTSVDQTSREIREYRKVQMLSEAFRSDCHKLEIERTRL 349
 QY 357 DDFLKAETKDLDETCVSSVTPDLTNCILPSPTK-MPGAVQDFTLLIRQISNDFRDL 415
 Db 350 -----KTLPM-----VPEFKPNQPMFCIKGLVLRVTRNLRNKQ 385
 QY 416 TLIHGAECIMSMTIGF-LYRGHSIQLSFMDTALLFMICALPENVIIDVISCYS 473
 Db 386 VVIMRLVQMLMGFLIFTLIRVQNMNMGAVDRGLLYQLGATPYGMLNANVLPRM 445
 QY 474 ERAMLTYELEDGLYTTGPFYFAKILGELPEHCAYIIYIGPTVLANLPAGLOPFLHL 533
 Db 446 LRAVSDQESQDGLQKQOMLAVYLAALPFSIVATVIFSSVCWTGLVPEVARF----- 500
 QY 534 LVWLVVFCRRIMALAAALLPTFHMAFFSNAL-----YNSPYLAGG 575
 Db 501 -----GYFSAALLAPHLIGEFLLTVILGAVQNPVINSYALLISGLDLSG 548
 QY 576 FMINSSLMTVPAMISKVSFLRMCCEGLMKIQF 608
 Db 549 FTRNDEMPPLKILGFTFOKICRILVNEP 581
 RESULT 9
 AA096990
 ID AA096990 standard; Protein; 651 AA.
 AC AA096990;
 XX 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R389H protein sequence.
 KM Human; ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutlein.
 XX Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 389 /note= "Wild-type Arg substituted by His"
 XX MO200227016-A2.

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX

Sequence 651 AA;

Query Match 19.9%; Score 697; DB 23; Length 651;
 Best Local Similarity 28.9%; Pred. No. 5,9e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

16 TPQDTSGIDRLFSSESNSLYFTYSGOPTLEVRDNLNVOYDLASQVPMFEOAOFKMPW 75
 8 TPGGSMGLQVNRGSSGLEGAPAT-APERSIGILHASTYSVSHRVR-PMWD-ITSCROOW 64
 76 TSPGQNSCELGIONLSFKVRSQOMLAIIGSSCGRASLDVITGR-GHGKIRKSGQIWI 134
 65 TROI-----LKDVSILYVESGOIMCIGSSGCKTLLDMSGRIGRAGTF-LGEYVY 115
 135 NGPSSPOLVRKCYAHVROHNOILNLTRETIATFADMRPRFSQAROKRVEDVYAE 194
 116 NGRARRKQFQDCSYVQSTPLSLTLVRETLHTALAI-RRGNPSPFKVEAVVAE 174
 195 LRLKQADTRVGNMVRNGLSGERRRVSIGVOLLMPGILLDEPTSGLDSTFHNLYKT 254
 175 LLSHVADRLNGISLIGISGERRRVSIAGLQDPVLMFDEPTTGLDCTANQIYVL 234
 255 LSLAKGNRLVLSLHOPRSDIFRLFDVLLMTSGTPIYIGAAOHMVOYFTAGVPCRY 314
 235 LVELARRRNVYVLIHOPRSELFDLFDIALISFCGLFCFPAEMLDFMDCGCPREH 294
 315 SNPDFYVDLSTIRSRREOLATREKASIALF-----LEKRDLDPLMK 362
 295 SNPDFYVDLSTIRSRREOLATREKASIALF-----LEKRDLDPLMK 362
 363 AETKDLDEDCVSSVPLDNLCLPSPK-MGAVVOFTLLIRQISNDFDLPTLLIHG 421
 349 -----KTLPM-----VPRKDSPGVFSKLGVLRRVTRNLVANKLAVITRL 390
 422 AEACIASMTIGFLYFG-----HSTQLSEMDTALLFMIGALIPFNVLIDVSKYSER 475
 391 LQNLIMGLFLFVLRRVSNVNLKGAIO-----DRVGLLYQFVGATPYGMLNVAVLPVLR 446
 476 AMLYELEDGLYTTGPFPAKIIIGELPEHCAYIIITGMPTWLANLRGLOPELLHLYV 535
 447 AVSQOESQDGLYQKQOMLALVLPSPVAVIMFSSVCVGTGLPEVARF-----499
 536 WLVECCRIMALAALLPFFHMASFVS-----NALYNSFYLAG-----GFM 577
 500 -----GYFAALLAPHLIGEFLLVLLGIVQNPINYSVALLSTAGVVGSGFL 549
 578 INLSLMTVPAWIKSVFLKRCFGLMKIQPSRRTYMPGLNLTIVAS 625
 550 RNIDEMPIFKRISYFTFOKCSILVAVNEFYGLNFCGSSNVSVTN 597

RESULT 11

AAU96993 standard; Protein; 651 AA.

AAU96993;

30-JUL-2002 (first entry)

XX Human ABCG5 mutant R419p protein sequence.
 DE Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 XX arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutelin.
 XX Homo sapiens.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 419
 FT /note="Wild-type Arg substituted by Pro"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI: 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 PS Claim 10; Page -: 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette
 gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 predisposition for developing sitosterolemia, arteriosclerosis or heart
 disease. The molecules of the invention are also useful for identifying
 a compound which alters ABCG5 activity level comprising contacting a cell
 culture or mammal with ABCG5 polypeptide with a compound and
 measuring ABCG5 biological activity in the cell culture or in mammal,
 where an increase or decrease in ABCG5 biological activity compared to
 ABCG5 biological activity in a control cell culture or mammal not
 contacted with the compound, identifies a compound that increases or
 decreases ABCG5 activity respectively. The cell culture or mammal
 comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 polypeptide in a cell culture or mammal is also compared with that of a
 second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419p protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX

Sequence 651 AA;

Query Match 19.9%; Score 697; DB 23; Length 651;
 Best Local Similarity 28.9%; Pred. No. 5,9e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

16 TPQDTSGIDRLFSSESNSLYFTYSGOPTLEVRDNLNVOYDLASQVPMFEOAOFKMPW 75
 8 TPGGSMGLQVNRGSSGLEGAPAT-APERSIGILHASTYSVSHRVR-PMWD-ITSCROOW 64
 76 TSPGQNSCELGIONLSFKVRSQOMLAIIGSSCGRASLDVITGR-GHGKIRKSGQIWI 134

Db 65 TRRI-----LKQSVLYVESQIMCIIAGSSSGKTTLLDAMSGLRAGTF--LEEYV 115

QY 135 NGQPSPPQLVKRCVAHYAHQHNOLLPLNLTARETLAFTLAOMRLPTFESOAGKREVEDYIAE 194

Db 116 NGRLARREGPODCEFSYVLOSIDLTLSSLTVAETLHTYALALAI--RQNGPSQKKEAVEAMAE 174

QY 195 LRLRCQADPRVGMNMYRGLSGGERRRVSIGCVOLLNMPGIIILDEPTSGDLSFTAHNLVKT 254

Db 175 LSLSHVADRILIGNYSLGGISGTGERRRVSIAOOLDPKVALLEDEPTSGDLCMTANOIVVL 234

QY 255 LSLRAGNRLVLSLHQPSRSDIRLFDVLNLTSGPIYIGAQNHVQYFTALGYCPRK 314

Db 235 LVELARNRNIIVLTITQPSSELFQDLKAITLSFGELIFCGYPAEMLDFFNDCGYCPPEH 294

QY 315 SNPADYVDLTJSDRRSRREQELATREKAOSIALP-----LEKRDLDQFLWK 362

Db 295 SNPFDEYMDLTSVDQSKREIEITSKRVQMIESAYKKSALCHTKLNIERMKHL----- 348

QY 363 AETKLDDETCVEBSSTPLDTNCLPSPTR--MEGAVQOFTLLIRQISNDFRDLPTLLIHG 421

Db 349 -----KTLPM-----VPRKTDSPGVESFKLVLLRLRYRNLRNKLAVITPL 390

QY 422 AEACLMSMTIGFLYEG-----HGSLOFPMQTAALLFMGALIPFNVLVDYSKYSER 475

Db 391 LQNLINGLELFFFLVLRVSNVLKGAQO----DPVGLLYQFAGATPYTGMLNVLNLEPVLR 446

QY 476 AMLYLELEDGLYTTGPPYFAKILGELPEHCAYIIYGMPTWMLANLRPGLOPFLHFLV 535

Db 447 AVSDQESQDGLYQKQOMQMLAYALHVLPEFVATMIFSSCYMTGLHPEYAR----- 499

QY 536 WLIVCCRIMALAAALPFTFHMAFSF-----NALNSFYLAG-----GPM 577

Db 500 -----GYESAALLAPHLIGEFELTVLGIYQNPRIANSVALLSIAGVLQSGSL 549

QY 578 IMLSLMTVPANISKVSFLRMCFEGIMKIQFSRRITKMLGLNLTAVS 625

Db 550 RNIQEMPIPEKTIISTFTFOKYCELIIVNEFYGLNFTCGSSNVSVTN 597

RESULT 12

AAEL3290

ID AAEL3290 standard; Protein; 651 AA.

XX AAEL3290;

AC

DT

XX

DE

Human sitosterolemia susceptibility gene (SSG) protein.

XX

KW Human; sitosterolemia susceptibility gene; SSG; atherosclerosis; sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy; gall stone; coronary heart disease; cardiovascular disease; arthritis; xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.

XX

SS Homo sapiens.

PN WO200179272-A2.

PD

XX

25-OCT-2001.

PE

18-APR-2001; 2001WO-US12758.

PR

18-APR-2000; 2000US-198465P.

PR

15-MAY-2000; 2000US-204234P.

PA (TULA-) TULARIK INC.

XX

XX Tlan H, Schultz J, Shan B;

FI

XX WPI; 2002-017598/02.

DR

N-PSDB; AAD22009.

XX

XX

Novel sitosterolemia susceptibility gene polypeptide and

RESULT 13
AAE31704
ID AAE31704 standard; Protein: 651 AA.
XX
AC AAE31704;
XX
DE 24-MAR-2003 (first entry)
XX
DE Human ABCG5 protein.
XX
KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW sitosterolemia; hyperlipidaemia; hypercholesterolemia; gall stone;
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW human; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
KW ABCG5.
XX
OS Homo sapiens.
XX
PN WO200281691-A2.
XX
PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-USA3823.
XX
PR 20-NOV-2000; 2000US-252235P.
PR 28-NOV-2000; 2000US-253645P.
XX
PA (TULAR) TULARIK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tlan H;
XX
DR WPI: 2003-058548/05.
DR N-PDB: AAD48882.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating
PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies -
XX
PS Claim 28; Page 78-79; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
CC invention are useful for treating or preventing sterol-related disorders
CC such as sitosterolemia, hyperlipidaemia, hypercholesterolemia, gall
CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC They are also useful in gene therapy. The present sequence is human
CC ABCG5 protein.
XX
SQ Sequence 651 AA;
Query Match 19.9%; Score 697; DB 24; Length 651;
Best Local Similarity 28.9%; Pred. No. 5.9e-62;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;
OY 16 TPQDTSGIDRLTFSSESNSLYFTYSGQPTLEVRDNTQYVDLASQYPRPQALQAFMPW 75
DB 8 TPGSGMGLQVNRGSSSLEGAAPAT-APEPHSLGILHASYSSSHVR-PWMD-ITSCROOW 64
OY 76 TSPSCONSCIEGLQNLSPKVASGOMLAISSGCGRASLDVITGR-GHGGKIKSGQIMI 134
DB 65 TROI-----LKQDVLYVESGQIMCIISSSGSKTLLDAMSGRIGRGCTF-LGEYIV 115
OY 135 NGQPSPOLYRKCAVAVRQHNOLLPLVTRETLAFLIAOMRLPTFSQAQRDRVEDYIAE 194
DB 116 NGRALRRRQPDQCSYVLOSPTLLSLVRETLHYTALLAI-RRGNPGSFOKKEAVMAE 174
OY 195 LRLHQCADTIRGNMYVAGLSGGERRRYSIGVQLMNGCILLDEPTSGIDSFTHNLVKT 254

DB 175 LSLSHVADRLIGNSLGIGISTGERRRVSIAQLLQDPRVMLFDEPTTGLCOMTANQIVLV 234
OY 255 LSRILAKGNRLVYLILSHQPRSDIFRLFDVILMTSGTPIYVGAQOHMYQVTAIGPCPRY 314
DB 235 LVELARRNRRIYVLIHQPRSELFDLPKIALISGELIFCGTPAEMLDPRDCGCPCEH 294
OY 315 SNPADFYVDLTSIDRSRECELAAREKAQSLALE-----LEKVRDLDDPLMK 362
DB 295 SNPDFYVDLTSVDQSKEREIETSKRQMESAYKKSAICHKTLKNIERKHL----- 348
OY 363 AETKDLDEDCVSESVPLDTNCLPSPFK-MPGAVQOFTLLIRQISNDFRDLPTLLIHG 421
DB 349 -----KTLPM-----VPFKTKDSPGVFSKIGVLLRRVTNLRVNRKLAIVITRL 390
OY 422 AEACLMSMTIGFLYFG-----HGSILSFMDTALIMGALIEPVNIIIDVSKCSER 475
DB 391 LQNLIMGLFLFLFVLRARSNVLKRAIQ-----DRGLIKYQVATPYGMALNAVLPVLR 446
OY 476 AMLYELEDGLYTTGPYFEAKILGELPEHCAYIITYGMPYWLNLNRGLQPELLRPLV 535
DB 447 AVSDQESQDGLYQKQWMLAYALHVLFPFSVATWIFSSVCYMTLGLHPEVARF----- 499
OY 536 WLIVYFCCRIMAAALLPFFHNASFS-----NALYNSFYLAG-----GFM 577
DB 500 -----GYFSALLAPHLIGELFLVLLGIQVQNPNIYNSVALLIAGVVGSGFL 549
OY 578 INLSLMTVPAMISKVSFLRMCEFLMKIKGFSRRTYMPGLNLIANS 625
DB 550 RNIOEMPIPRKIISTYFTFOKCSILVNEBYGILNFGSSNVSVTIN 597
RESULT 14
AAU96989
ID AAU96989 standard; Protein: 651 AA.
XX
AC AAU96989;
XX
DE 30-JUL-2002 (first entry)
XX
DE Human ABCG5 mutant R419H protein sequence.
XX
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 419
FT /note= "Wild-type Arg substituted by His"
XX
PN WO200227016-A2.
XX
PD 04-APR-2002.
XX
PF 25-SEP-2001; 2001WO-US29859.
XX
PR 25-SEP-2000; 2000US-235268P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
PI Patel SB, Dean M;
XX
DR WPI: 2002-416483/44.
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, atherosclerosis and heart diseases -
XX

PS Claim 9; Page -: 66pp; English.

XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal.
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant R419H protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AAU96984) given on pages 35-36
CC of the specification.

XX Sequence 651 AA:

Query Match 19.9%; Score 696; DB 23; Length 651;
Best Local Similarity 28.9%; Pred. No. 7.5e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

QY 16 TPQDPSGIDRLFFSESDNSLFTYSGQNTLEVDLNTQVDLASQWPEQDLQAFKMPW 75
DB 8 TPGGSMGLQVNRGSSSLGAPAT-APEPHSIGILHASYSVSHRRV-PWMD-ITSCROQM 64
QY 76 TSPSCONSCELGIONMSPFVRSOGMLATIGSSGGRASLDVITGR-GHGCKIKSGQITV 134
DB 65 TROI-----LKVSLTVESGQIMCLIGSSGSKTLLDMSRILGAGTF-LGEYVV 115
QY 135 NGOPSSPOLVRKCVAVRHQNDLNLVRETIAFIAQMRLEPTFSQAQRDRVEDVIAE 194
DB 116 NGRALRRQFOQCFYVQSDTLSSLVRETLHTALLAT-RRGNPSSFOKVAEVAWE 174
QY 195 LRLQCAATRVGNMTVVRGLSGERRRVSIGVLLMNPGLILDEPTSGIDSTFAHNLVKT 254
DB 175 LLSHVAADRLIGNVSLGISTGERRRVISAQLLQDPKMLFDEPTTGIDCMTANQIVYL 234
QY 255 LSLAKGNRLVLIHOPRSDIFRFLDVLMTSTPTIYCAQAQMVOTFRAIGPCPRY 314
DB 235 LVELARRRRIYVLTTHOPRSELEFOLFEDKIALISFELIFCGPAPMLDFFNCGGCPCEH 294
QY 315 SNPAFDYVLTSDIRSRHEQELATREKAQSLAALF-----LEKVRDLDFLMK 362
DB 295 SNPFDFYDLTSVDQSKEREIETSKRYQMIESAKKSAICKTKIETKRMKHL----- 348
QY 363 AETKDLDEDTCEVSSVTLDNCLFSPPK-MPGAVQPTTLIRQISNDFDLPTLLHG 421
DB 349 -----KTLFPM-----VPEKTKDSGFYSKLGVLIRVTRMLVRKRLAVITRL 390
QY 422 AEACLSMTIGFLYFG-----HGSIOISFMDTALLEMICALIFENVIDVISCYSEK 475
DB 391 LQNLINGLFLFVLVRANSVNLKGAIQ-----DHVGLAQFGAPPYTGMNAVNLEPVLK 446
QY 476 AMLYELEDGLYTTGPFFAKILGELPEHCAYIIITYGMPYTLANLRGLOPFLHFLV 535
DB 447 AVSDQESODGLYQKQWMLAVLAHLVLPESVATMIFSSVCWTICLHEVARF----- 499
QY 536 WLAVYCCRIALAAALPTFHASFS-----NALYNSFYLAG-----GFM 577
DB 500 -----GYFSAALLAPHLIGEFLLVLLGIVQNPNIYNSVALLSTINGVLGSGFL 549

QY 578 INLSIMTVPAWISKVSEFLMCEGLMKIOFSRRTRYKMPIGNLTAVS 625
DB 550 RNIQEMPIRKIIISYFTFOKVCSEILVNEFGLNFCGSSNVSVTTN 597
RESULT 15
ID AAU96992 standard; Protein: 651 AA.
XX AAU96992:
AC
XX
DT 30-JUL-2002 (first entry)
XX
DE Human ABCG5 mutant E146Q protein sequence.
XX
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KM mutant; mutcin.
XX
OS Homo sapiens.
OS Synthetic.
OS
FH Key Location/Qualifiers
FT Misc-difference 146 /note="Wild-type Glu substituted by Gln"
FT
XX
PN W0200227016-A2.
XX
PD 04-APR-2002.
XX
PE 25-SEP-2001; 2001WO-US29859.
XX
PR 25-SEP-2000; 2000US-235268P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
PI Patel SB, Dean M;
PI
DR MPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
PS
XX
XX Claim 12; Page -: 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal.
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant E146Q protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AAU96984) given on pages 35-36
CC of the specification.

50 Sequence 651 AA;

Query Match	Score	DB	Length
19.88;	694;	23;	651;
Best Local Similarity	39.79;	Prod No. 1	29.61;

Best Local Similarity 28.7%; Pred. No. 1.2e-61;
Matches 186; Conservative 125; Mismatches 241; Indels 96; Gaps 16;

```

OY 16 TPQDPSGJODRLFSESJDSLSLYTSGQNTLEVRJLNOYDLASQVPPREQLAQRKMP 75
Db 8 TPGSGMGJQVNRGSSSLLEGAPATAPEPHSJGILHASTVSXHRVPMWDITSCROOW 64
OY 76 TSPSCNSCEJGIONLSEKVRNSGOMALAISSGCCGASJLDYVTGRGAGKIKSGQIWI 134
Db 65 TRQI-----LKDVSLYVESGQIMCJLSSGSGKTTLLDMSGRJGAGTFJDEYVY 115
OY 135 NGQPSRPOLYARKCAVHVRQHNDLNLVYREJLAFIAQRKLRPTESQAOBRKVEDYIAE 194
Db 116 NGRALRRQFQDCFSYVLQSDJLSSLYVROTJLHTYATLAI-RQNDGSEFOKVEEYVME 174
OY 195 LRLRQCADTRVGNMVRGLSGGERRRVSIQVOLLNMPJGLIIDEPTSGDSFAHMLVMT 254
Db 175 LSLSHVADRLIGNSLGGSTGBERRRVSIAADLDDPKMLDEPTJGLDCMTANQIVVL 234
OY 255 LSLRAKGRNLVLISLHQRSDIFRLFDVLVLLMTSGTPIYIGAQHMYOFTAGYPCPRY 314
Db 235 LVEJLARRRNRIVYLTHQRESELFOPEJKAJLISFGLICFPAPKXLDPEFNDGCGYCPBH 294
OY 315 SNPDADFYVDJNSIDRRSREQELATEKQOSLALF-----LEKVRJLDJFLMK 362
Db 295 SNPDFYJMDLVSVDQSKERELETERKRVOMISAKKSAICRKTJKNIERMKL----- 348
OY 363 AETKDIDEDTCVESSVTPLDNTCNLESPTK-MEQAVOQFETLLIRQJNSNFDRLPTELLHG 421
Db 349 -----KTLPM-----VPKFKTRDSGVSCKGLVLLRRVTRNLVRKRLAVITRL 390
OY 422 AEACJLMSFTIGFLYG-----HGSIQJSMFDPALLENIGALIPBNVILLDIYSKCYSER 475
Db 391 LQNLIMGJLLEFVLRRVNSLKGAIQ-----DRVGLLJOFGATPYTGKMLNVLNPEVLR 446
OY 476 AMLYEJLEDDGLYTTGPYPFPAKJLIGELPEHCAYIIYGMPTYMLANLRPELOPFLHFLV 535
Db 447 AVSGQESDQJGLQKQOMMLAYALHVLPPSVATMTFFSVCYTTJGLIHPVARR----- 499
OY 536 WLVPVCCRIMALAALALPETHMASFS-----NALYNSFYLAG-----GFM 577
Db 500 -----GYFSALLAPHLIGELFLVTLVJLQVNPRIYNSVALLSIAGVLGSGPL 549
OY 578 INLSJLMTVPAMISVSLFRCQCEJMLKIOPSRRTYKMPGLJLTAVS 625
Db 550 RNIDMPLPEKIIISFTOKYCSSELVYNEFGLNFGTCCSSSVATYN 597

```

Search completed: July 25, 2003, 17:12:13
Job time : 90 secs

us-09-989-981a-7.rng

Page 1

OH nucleic - nucleic search, using sw model

Run on: July 26, 2003, 17:02:14 ; Search time 645.803 Seconds

(Without alignments) :
1156.341 Million cell updates/sec

Title:	US-09-989-981A-7
Perfect score:	2669
Sequence:	1 ggtgcctgtctcaggaac.....caattaaatgtattgagc 2669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03: *

1:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25:	/SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Description			
No.	Score	Match	Length	ID	
1	2669	100.0	2669	25	AAD48883
2	1680.6	63.0	3239	24	AAK83218
3	1499	56.2	2564	24	ABN90022
4	1430	53.6	2019	25	AAD48881
5	291.6	10.9	580	22	AAH95911
6	203.6	7.6	1920	24	ABK51681
7	203.6	7.6	2340	24	AAD22009
8	203.6	7.6	2340	25	AAD48882

[illegible]

ALIGNMENTS

RESULT 1	
AAD48883	
ID	AAD48883 standard; DNA; 2669 bp.
XX	
AC	AAD48883;
XX	
DT	24-MAR-2003 (first entry)
XX	
DE	
XX	
XX	Human ABCG8 DNA.
KW	ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW	sitosterolaemia; hyperlipidemia; hypercholesterolaemia; gall stone;
KW	HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW	human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
ABCG8;	gene; ds.

Homo sapiens.

Key	Location/Qualifiers
CDS	100 3131

```

100..2121
/*tag= a
/product= "hABCG8 protein"

```

WO200281691-A2

17-OCT-2002.

20-NOV-2001; 2001WO-US43823.

PR 20-NOV-2000; 2000US-252235P.
PR 28-NOV-2000; 2000US-253645P.

XX (TULAR) TULARIK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
XX MPI: 2003-058548/05.
DR P-PSDB: AAE31705.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating
PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies -
XX
PS Claim 13; Page 80; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
CC invention are useful for treating or preventing sterol-related disorders
CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC They are also useful in gene therapy. The present sequence is human
CC ABCG8 DNA.
XX
SQ Sequence 2669 bp; 595 A; 768 C; 722 G; 584 T; 0 other:

Query Match 100.0%; Score:2669; DB 25; Length 2669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGTCCCTGCTCCAGAAACAGAGTGAAGACATGCGCCCTGGCAGGACAGCTGGTCT 60
DB 1 GTGTCCCTGCTCCAGAAACAGAGTGAAGACATGCGCCCTGGCAGGACAGCTGGTCT 60
OY 61 AAGAGAGTGCAGCCCGAGGCTCAGACAGCTGGGCGCCGATGCGCGGGAAGCGGAG 120
DB 61 AAGAGAGTGCAGCCCGAGGCTCAGACAGCTGGGCGCCGATGCGCGGGAAGCGGAG 120
OY 121 GAGAGAGGCTGCCGAAAGGGGCGACCTCCAGAGTACCTGGGCGCTCCAGATAGATTG 180
DB 121 GAGAGAGGCTGCCGAAAGGGGCGACCTCCAGAGTACCTGGGCGCTCCAGATAGATTG 180
OY 181 TTCCTCTGTAAGTGAACAGAGCTGACTTCACTAAGTGGCCAGCCCAACACCCTG 240
DB 181 TTCCTCTGTAAGTGAACAGAGCTGACTTCACTAAGTGGCCAGCCCAACACCCTG 240
OY 241 GAGGTGAGAGACCTCACTACAGAGTGGAGTGGGCTGAGTGGCTGCTGGTGGAGCAG 300
DB 241 GAGGTGAGAGACCTCACTACAGAGTGGAGTGGGCTGAGTGGCTGCTGGTGGAGCAG 300
OY 301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGCCAGATTTCTTGAGCTGGGC 360
DB 301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGCCAGATTTCTTGAGCTGGGC 360
OY 361 ATCCAGAACCTTAAGCTTCAAGTGAAGTGGGAGATGCTGGCCATATAGGAGCTCA 420
DB 361 ATCCAGAACCTTAAGCTTCAAGTGAAGTGGGAGATGCTGGCCATATAGGAGCTCA 420
OY 421 GGTGTGGAGAGGCTCTGCTAGATGTATCATCTGGGCGGAGGTCAAGGCGCAATC 480
DB 421 GGTGTGGAGAGGCTCTGCTAGATGTATCATCTGGGCGGAGGTCAAGGCGCAATC 480
OY 481 AAGTCAGGCCAGATCTGGATCAATGGGCGAGCCGCTCGCTCAGCTGGTGAAGAAATGT 540
DB 481 AAGTCAGGCCAGATCTGGATCAATGGGCGAGCCGCTCGCTCAGCTGGTGAAGAAATGT 540
OY 541 GTGGCCACAGTGGGCGGACACACAGCTGCTCCCAACTGATGTGCGCAGAGAACCTTG 600
DB 541 GTGGCCACAGTGGGCGGACACACAGCTGCTCCCAACTGATGTGCGCAGAGAACCTTG 600
OY 601 GCCTTCATGCTCCAGATGGGCTGGCCAGAAACCTTCTCCAGGCGCCAGGCGACAAAG 660

DB 601 GCCTTCATGCTCCAGATGGGCTGGCCAGAAACCTTCTCCAGGCGCCAGGCGACAAAG 660
OY 661 GTGAGAGAGTATGGGAGGCTGGGCTTGGAGCTGGGCTGACACACCGGCTGGGCAAC 720
DB 661 GTGAGAGAGTATGGGAGGCTGGGCTTGGAGCTGGGCTGACACACCGGCTGGGCAAC 720
OY 721 ATGTACGTGGGCGGCTTCTGGGCGGAGAGGAGAGTATGATGATGGGCTGAGCTC 780
DB 721 ATGTACGTGGGCGGCTTCTGGGCGGAGAGGAGAGTATGATGATGGGCTGAGCTC 780
OY 781 CTGTGGAACCCAGGAATCTTATCTGACAGAACCCACTTGGGCTGAGAGCTTACA 840
DB 781 CTGTGGAACCCAGGAATCTTATCTGACAGAACCCACTTGGGCTGAGAGCTTACA 840
OY 841 GCCCAGACCTGGTGAAGACCTTGTCCAGGCTGGGCGCAAGGCAACCGGCTGGTCAAC 900
DB 841 GCCCAGACCTGGTGAAGACCTTGTCCAGGCTGGGCGCAAGGCAACCGGCTGGTCAAC 900
OY 901 TCCTTCACACAGCTCGCTGACATCTTACAGGCTGTTGATGCTGCTCTGATGACG 960
DB 901 TCCTTCACACAGCTCGCTGACATCTTACAGGCTGTTGATGCTGCTCTGATGACG 960
OY 961 TCTGGACCCCATCTACTAGAGGCGGCGCCAGACATGCTCAGATTTTCAAGCCATC 1020
DB 961 TCTGGACCCCATCTACTAGAGGCGGCGCCAGACATGCTCAGATTTTCAAGCCATC 1020
OY 1021 GGTACCCCTGCTCTGCTACAGCAATCTGCTGATCTATGTGAGCTTACCAAGATT 1080
DB 1021 GGTACCCCTGCTCTGCTACAGCAATCTGCTGATCTATGTGAGCTTACCAAGATT 1080
OY 1081 GACAGGCGCAGCAGAGAGAGAGATTTGGCCACCGAGGAGAGGCTCAGTCACTGGCAGCC 1140
DB 1081 GACAGGCGCAGCAGAGAGAGAGATTTGGCCACCGAGGAGAGGCTCAGTCACTGGCAGCC 1140
OY 1141 CTGTTTCTAGAAAAGTGTGCTGACTTATGATGACTTTTATGGAAGAGAGACAGAGAT 1200
DB 1141 CTGTTTCTAGAAAAGTGTGCTGACTTATGATGACTTTTATGGAAGAGAGACAGAGAT 1200
OY 1201 CTTGACGAGGACCTGTGTGGAAGAGAGCTGACCCCACTAGACACCAACTGCTCCG 1260
DB 1201 CTTGACGAGGACCTGTGTGGAAGAGAGCTGACCCCACTAGACACCAACTGCTCCG 1260
OY 1261 AGTCTACAGAGATGCTTGGGCGGCTGACAGAGTTCAGAGCTGATCCGTCGACAGATT 1320
DB 1261 AGTCTACAGAGATGCTTGGGCGGCTGACAGAGTTCAGAGCTGATCCGTCGACAGATT 1320
OY 1321 TCACAGACATTCGAGACCTGACCCTCTCATTCATGATGGGCGGAGGCTGTCTGATG 1380
DB 1321 TCACAGACATTCGAGACCTGACCCTCTCATTCATGATGGGCGGAGGCTGTCTGATG 1380
OY 1381 TCATGACATCGGCTCTCTATTTTGGGCAATGGGAGCAATCCAGCTCTCTCATATGAT 1440
DB 1381 TCATGACATCGGCTCTCTATTTTGGGCAATGGGAGCAATCCAGCTCTCTCATATGAT 1440
OY 1441 ACAGCCGCCCTCTTGTTCATGATCGTGTCTCATCCCTTCAACGCTATTTGATGTC 1500
DB 1441 ACAGCCGCCCTCTTGTTCATGATCGTGTCTCATCCCTTCAACGCTATTTGATGTC 1500
OY 1501 ATCTCCAAATGTTACTAGAGAGGCGCAATCTTACTATGAACTGGAAGACGGCTGTAC 1560
DB 1501 ATCTCCAAATGTTACTAGAGAGGCGCAATCTTACTATGAACTGGAAGACGGCTGTAC 1560
OY 1561 ACCACGTGTCATATTTCTTGGCAAGATCTCTGGGAGAGCTTCCGAGGACCTGTGCTAC 1620
DB 1561 ACCACGTGTCATATTTCTTGGCAAGATCTCTGGGAGAGCTTCCGAGGACCTGTGCTAC 1620
OY 1621 ATCATCATCTAGAGGATGCCACCTACTGCTGGCCAACTGAGAGGCGGCTGACAGCCC 1680
DB 1621 ATCATCATCTAGAGGATGCCACCTACTGCTGGCCAACTGAGAGGCGGCTGACAGCCC 1680
OY 1681 TTCCTGTGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

Dp	1681	TTCCGCTGCACACTTCCTCCTGTGGTGGGTGGTGCTTCTCTTCAGAGATTATGACCCTG	1740
Oy	1741	GCGCGCGGGGCCCTGCTCTCCCACTCCACAATGSGCTCTCTTCTTCAGCAATGCCCTCTAC	1800
Dp	1741	GCCCCGGGGGCCCTGTCTCCCACTCCACAATGSGCTCTCTTCTTCAGCAATGCCCTCTAC	1800
Oy	1801	AACCTCTCTTAACCTCTCGCGGGGGGGCTCATATGATAAAGTGAACCTGAGCACCTGTGGACAGTCCC	1860
Dp	1801	AACCTCTCTTAACCTCTCGCGGGGGGGCTCATATGATAAAGTGAACCTGAGCACCTGTGGACAGTCCC	1860
Oy	1861	GCGTGGATTTCCAAAGTGTCTCTCTCTCGGGTGGTGTGTTTTGAAGGGCTGATGAAGAATTACG	1920
Dp	1861	GCGTGGATTTCCAAAGTGTCTCTCTCTCGGGTGGTGTGTTTTGAAGGGCTGATGAAGAATTACG	1920
Oy	1921	TTCAGCAGACACTTATATAAATCCTCTCGGGAACTCACCATCGGGGTCTCAGAGAT	1980
Dp	1921	TTCAGCAGACACTTATATAAATCCTCTCGGGAACTCACCATCGGGGTCTCAGAGAT	1980
Oy	1981	AAAAATCCATGSCATGAGAGACTGAGACTGAGACTGACCTGACCTCTAACGGCATCTACCTCATGTC	2040
Dp	1981	AAAAATCCATGSCATGAGAGACTGAGACTGAGACTGACCTGACCTCTAACGGCATCTACCTCATGTC	2040
Oy	2041	ATTGGCCTCAGCGGGTGGCTTCATGATGGTCCCTGTACTACAGTGTCTTAAAGTTTCATCAACAG	2100
Dp	2041	ATTGGCCTCAGCGGGTGGCTTCATGATGGTCCCTGTACTACAGTGTCTTAAAGTTTCATCAACAG	2100
Oy	2101	AAACCAAGTCAAGACTGTGTATTCAGAGCCAGACAGCTGTCCCGGTGTGGGGAGACTGTAGC	2160
Dp	2101	AAACCAAGTCAAGACTGTGTATTCAGAGCCAGACAGCTGTCCCGGTGTGGGGAGACTGTAGC	2160
Oy	2161	AGACCCCTTCAACTGCACTCTCCCTCTCTCAGAGAGCCCTTCTCTGGGAGCAGTGAAGCAATGA	2220
Dp	2161	AGACCCCTTCAACTGCACTCTCCCTCTCTCAGAGAGCCCTTCTCTGGGAGCAGTGAAGCAATGA	2220
Oy	2221	CCCTACAGAGTCTCAGCTACATCCCGGGCCAGAGGGTGGCTGAGTGGCCANAGACCAGCCACAG	2280
Dp	2221	CCCTACAGAGTCTCAGCTACATCCCGGGCCAGAGGGTGGCTGAGTGGCCANAGACCAGCCACAG	2280
Oy	2281	GATGGCAGTAGAATAAAGACAGTGGAAAGGAGATTTCGTCTCAGTGGCAGGAGAGACTCGAT	2340
Dp	2281	GATGGCAGTAGAATAAAGACAGTGGAAAGGAGATTTCGTCTCAGTGGCAGGAGAGACTCGAT	2340
Oy	2341	GACGTGGAGAAAACCTGCACTCGGTGGCAGCTCAACAGTGTGAATTATTATTCCTTTTGA	2400
Dp	2341	GACGTGGAGAAAACCTGCACTCGGTGGCAGCTCAACAGTGTGAATTATTATTCCTTTTGA	2400
Oy	2401	TATGCAATTTATATAGGCAATCGATATAGATGAGGAGCAAACTAAGAAATTAATTTGGGTAG	2460
Dp	2401	TATGCAATTTATATAGGCAATCGATATAGATGAGGAGCAAACTAAGAAATTAATTTGGGTAG	2460
Oy	2461	CTAGACTGTGCAGGAATTTGTGGAACCTGTGAGGGAACAATAAAGTAGTAGCAGATTGG	2520
Dp	2461	CTAGACTGTGCAGGAATTTGTGGAACCTGTGAGGGAACAATAAAGTAGTAGCAGATTGG	2520
Oy	2521	GCTTCATCTTCCAGGGGGCCCCACACTCCGTGTGTAGCCACCATCAATTACAGAAAGTAGCC	2580
Dp	2521	GCTTCATCTTCCAGGGGGCCCCACACTCCGTGTGTAGCCACCATCAATTACAGAAAGTAGCC	2580
Oy	2581	TAAAGATTTACACAGAAGATGCCATGCCCTCTTTTGTGTGGGGTCAATGGGCTCCAAAAGC	2640
Dp	2581	TAAAGATTTACACAGAAGATGCCATGCCCTCTTTTGTGTGGGGTCAATGGGCTCCAAAAGC	2640
Oy	2641	CAACGTGAACAATTTAAATAATGTATTGAGC 2669	
Dp	2641	CAACGTGAACAATTTAAATAATGTATTGAGC 2669	

RESULT 2
ABR83218 standard; cDNA; 3239 BP.

XX ABR83218;
XX

27-AUG-2002 (first entry)

Human transporter and ion channel, TRICH9, Incyte ID 6585710CB1, CDNA.

Human, ss; gene; transporter and ion channel; TRICH; transport disorder; neurological disorder; muscle disorder; immunological disorder; cancer; scleroderma; systemic lupus erythematosus; allergy; leukemia; cell proliferative disorder; cervical cancer; breast cancer; neurodegenerative disorder; Parkinson's disease; Alzheimer's disease; myotonic dystrophy; catatonia; endocrine disorder; diabetes; renal disorder; Good pasture's syndrome; viral, bacterial, fungal, parasitic, protozoal infection; Good pasture's syndrome; viral infection; cirrhosis; bacterial infection; fungal infection; parasitic infection; protozoal infection; helminthic infection; cardiovascular disorder; atherosclerosis; hepatic disease.

Homo sapiens.

WO200240541-A2.

23-MAY-2002.

25-OCT-2001; 2001WO-US46055.

27-OCT-2000; 2000US-243989P.

03-NOV-2000; 2000US-245904P.

09-NOV-2000; 2000US-247673P.

17-NOV-2000; 2000US-249661P.

20-NOV-2000; 2000US-252232P.

01-DEC-2000; 2000US-250790P.

(INCYTE-) INCYTE GENOMICS INC.

Tang YT, Yue H, Nguyen DB, Hafalia AYA, Elliott VS, Lu Y; Wallia NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjanmala M; Rankumam J, Arvilcu G, Gietzen KJ, Lall PG, Azimzal Y, Khan FA; Thangavelu K, Thornton M, Lu DM, Tribouley CM, Warren BA; Ison CH, Das D, Raumann BE, Policky JL, Kearney L; WPI; 2002:463570/49.

P-PSDB: ABG61539.

New transporters and ion channels (TRICH) polypeptides, useful for diagnosing, preventing, and treating disorders associated with an abnormal expression or activity of TRICH, e.g. immunological, muscular or renal disorders

Claim 5; Page 167-168; 178pp; English.

The invention relates to human transporters and ion channels (TRICH) polypeptides, a naturally occurring amino acid sequence 90 & identical to TRICH, a biologically active fragment of TRICH or an immunogenic fragment of TRICH. Also included are an isolated polynucleotide encoding TRICH, a recombinant polynucleotide comprising a promoter sequence operably linked to the TRICH polynucleotide, a cell transformed with the recombinant polynucleotide, a transgenic organism comprising the recombinant polynucleotide, an isolated antibody that binds specifically to TRICH, and screening for compounds which bind to TRICH, modulate TRICH, modulate TRICH expression or are ant/agonists of TRICH.

The polypeptides are useful for diagnosing, treating, and preventing transport, neurological, muscle, immunological disorders (e.g. scleroderma, systemic lupus erythematosus, allergies), cell proliferative disorders such as cancers (e.g. leukemia, cervical or breast cancers), neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes, Grave's disease), gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Good pasture's syndrome), viral, bacterial, fungal, parasitic, protozoal and helminthic infections, cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many other diseases and disorders detailed in the specification. They can also be used in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and

OY	460	CGAGGTCACGGGGGCAAGATCAAGTCAAGGCCAATCTGGATTCATATGGGCAACGCCAGCTCG	519
Db	361	AGAGGCCACGGTGGCAAGATGAAATGACGCAAAATTTGGATTAATGGGCAACCCAGTACG	420
OY	520	CCCTACCTGGTGAAGAAATGTGTGGCCCACTGTGCGGCAGCAACAACGAGCTGGTCCCAAC	579
Db	421	CCCTACCTGGTGAAGAAATGTGTGGCCCACTGTGCGGCAGCAACAACGAGCTGGTCCCAAC	480
OY	580	TTGACTGTGCGAGAGACCTTGGGCTTCATTGGCCAGATGGGCGCTGCCAGAACTCTTCC	639
Db	481	CTACACGTCAGAGAGACCCCTGGCTTCATTGGCCAGATGGGCGCTGCCAGAACTCTTCC	540
OY	640	CAGGCCAGGTTGACAAAAGGGGAGAGACGTATGCGGAGCTGGCGCTTAGGCACTGC	699
Db	541	CAGGCCAGGTTGACAAAAGGGGAGAGACGTATGCGGAGCTGGCGCTTAGGCACTGC	600
OY	700	GCTGACACCCGCGTGGGCAACAATGTATCGTGGGGGGTTGTGGGGGGGTAGGGCAGAGAA	759
Db	601	GCCAAACCAAGAGTGGGGCAACAAGTATGTACGTGGGGGTGTCCGGGGGTAGGGCAGAGAA	660
OY	760	GTACAGCTTGGGGTGCAGCTCCTGTGGAAACCAGAACTCTTATTTCTGACAGAACCCACC	819
Db	661	GTGAGCATTTGGGGTGCAGCTCCTGTGGAAACCAGAACTCTTATTTCTGATGAAOCCACT	720
OY	820	TCGGGGCTGAGACGCTTCACACCCCAACCTGTGTAAACCTTTGTCCAGGCTGGCCAAA	879
Db	721	TCGGGGCTGAGACGCTTCACACCCCAACCTGTGTAAACCTTTGTCCAGGCTGGCCAAA	780
OY	880	GGCAACCGGCTGTGCTCATCTCCCTCCACAGACCTCGCTGACATCTTCAAGGCTGT	939
Db	781	GGCAACCGGCTGTGCTCATCTCCCTCCACAGACCTCGCTGACATCTTCAAGGCTGT	840
OY	940	GATCTGTCTCCTCATATGACGCTGTGGCACCCCACTACTTAAAGGGGCGGCCAGCAATG	999
Db	841	GACTGTGTCTCCTCATATGACGCTGTGGCACCCCACTACTTAAAGGGGCGGCCAGCAATG	900
OY	1000	GTCGAGTATTTCACAGCCATCGGGTACCCTGTCTCGCTACAGCAATCTGTGACTTC	1055
Db	901	GTCGAGTATTTCACAGCCATCGGGTACCCTGTCTCGCTACAGCAATCTGTGACTTC	960
OY	1060	TATGTGACCTGTGACCAAGATTGACAGGGCAGACAGAGACAGAAATTTGGCCACAGGGAG	1115
Db	961	TACGTGACCTGTGACCAAGATTGACAGGGCAGACAGAGACAGAAATTTGGCCACAGGGAG	1020
OY	1120	AAGGCTCAGTCACTGCGACCCCTGTTTCTAGAAAAAGTGCCTGATCTAGATCACTTTCTA	1177
Db	1021	AAGGCTCAGTCACTGCGACCCCTGTTTCTAGAAAAAGTGCCTGATCTAGATCACTTTCTG	1080
OY	1180	TGGAAAGGAGAGAACGAAGATCTTACAGAGACACCTGTGTGGAAGACAGCGGTGACCCCA	1235
Db	1081	TGGAAAGGAGAGAACGAAGATCTTACAGAGACACCTGTGTGGAAGACAGCGGTGACCCCA	1140
OY	1240	CTAGACACCAACTGCGCTCCGAGTCTCTCGAAGATGCGTGGGGGGGTGCAGAGATTACG	1295
Db	1141	CAGAGCACTGACTG---TGGACTGTCTGTGAGCTGCCCGGAGTATGAGCAGATTTC	1197
OY	1300	ACGCTGATCGCTCGTCAATTTCCAAAGCACTTCGAGACCTGCCCCACCTCTCATCATCAT	1358
Db	1198	ACCGTGAATCGCTGCAATTTCCAAATGATCTCCGGGAGACGTGCCACAGCTGTCTATCAT	1257
OY	1360	GGGGGGGAGGCGTGTCTATATGTCATATGACATCGGCTCTCTATTTTGGCCATGGAGGC	1419
Db	1258	GGGGGGGAGGCGTGTCTATATGTCATATGACATCGGCTCTCTATTTTGGCCATGGAGGC	1317
OY	1420	ATCCAGCTCTCTTCATGATGATACAGCCGCCCTTGTTCATGATGCGGTCTATCCCT	1479
Db	1318	AAGCAGCTCTCTTCATGATGATGATACAGCCGCCCTCTTCATGATGATGCGGTCTATCCCT	1377
OY	1480	TTCAACGTCATTTGATGATCATCTCCAAATGTTACTACAGAGGAGCAATCTTTACTAT	1539
Db	1378	TTCAATGTCATCTGATGATGTGTCTCCAAATGTCACTCGAGAGGTCAATCTGTACTAT	1437
OY	1540	GAACTGGAAGACGGGCTGTACACACAGTGTCCATATTTCTTTGCCAAGATCTCGGGAG	1599

Db	1438	GAGCTGGAGAGAGGGGCTGTACACTGCTGGTCTTATTTCTTTGCCAAGATCCTAGACAGA	1497
Qy	1600	CTTCCGAGACACTGTGACCTACATCATCTACGGGATGCCACTACTGGCTGCCAAC	1655
Db	1498	TTGCCGGACACTGTGACCTACACTCATCTACGGGATGCCACTACTGGCTGCCAAC	1557
Qy	1660	CTGAGCCGAGGCTCCAGGCTTCTGCTGGACCTTCGATGGTGGTGGCTGGTGGCTTC	1719
Db	1558	CTGCGGCGCGGCTCCGAGGCTCTTCTTCTTACACTCTCTGCTGCTGCTGGTGGTGGCTTC	1617
Qy	1720	TGTTGACAGATTATGAGCCCTGGCGCGCGGCGGCTTCGCCACCTTCACATGGGCTTC	1779
Db	1618	TGCTGAGAGACCATGAGCCCTGGCGCTCTCTGCTGACATGCGCCACCTTCACATGGCTTC	1677
Qy	1780	TTCTTCAGCAATGAGCCCTTCACAACTCTTCATCTGAGGCGGGGCTTCATATTAACCTTG	1839
Db	1678	TTCTTCTGCAATGAGCCCTTCACAACTCTTCATCTGAGGCGGGGCTTCATATTAACCTTG	1737
Qy	1840	AGCAGCCTGTGACAGAGTCCCGCGCTGATTTCCAAAGTCTCTGCTGCTGGTGGTTC	1899
Db	1738	GACAACCTGTGATGATGCTGCTGCATGATGATCTCCAAAGCTGTGCTTCCTCCGCTGGCTTC	1797
Qy	1900	GAAAGGCTGTATCAAACTTCAGTTCAGCAGAGAACTTATAAATGCTCTCGGAACTTC	1959
Db	1798	TGCGGGCGTATCAAACTTCATTAATGAGCACTTTCACACCAAAATGCGCAACTTC	1857
Qy	1960	ACCATCTGCGGCTCTGAGAGATTAATTCCTAGTGCATGAGAGCTGAGACTGCTACCTCTC	2019
Db	1858	ACCTTCTCTCATCTGCTGAGACACGATGATCATGCTGAGACCTGAACTGCACTCACTC	1917
Qy	2020	TACGCCATCTACCTTCATGCTGATTTGGCCTCAGCGGTGCTTCATGCTCTGACTACGTG	2079
Db	1918	TATGCGATCTACCTTCATGCTGATTTGGCCTCAGCGGTGCTTCATGCTCTGACTACGTG	1977
Qy	2080	TGCTTAAGTTCATCAAAACAGAAACCAAGTCAAGCTGTGTA	2121
Db	1978	TGCTTAAGTTCATCAAAACAGAAAGTCAAGCTGTGTA	2019
RESULT 5			
AAH98911			
ID	AAH98911	standard; cDNA; 580 BP.	
AC	AAH98911;		
DT	12-OCT-2001	(first entry)	
DE	Arabidopsis EST-derived coding sequence SEQ ID NO: 768.		
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; magpie; horse;		
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;		
KW	diagnostics; forensic test; gene mapping; genetic disorder;		
KW	biodiversity; gene therapy; nutrition; ss.		
OS	Arabidopsis thaliana.		
XX	PM2020154477-A2.		
XX	02-AUG-2001.		
XX	25-JAN-2001; 2001WO-US02687.		
XX	25-JAN-2000; 2000US-0491404.		
PR	17-JUL-2000; 2000US-0617746.		
PR	03-AUG-2000; 2000US-0631451.		
PR	15-SEP-2000; 2000US-0663870.		
XX	(HYSE-) HYSEQ INC.		
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;		
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;		

DR WPI: 2001-476164/51.
DR P-PSDB: AAM24252.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX antibodies and research use -
XX
PS Claim 1: Page 664; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SO Sequence 580 BP; 146 A; 154 C; 116 G; 164 T; 0 other;

Query Match 10.9%; Score 291.6; DB 22; Length 580;
Best Local Similarity 97.1%; Pred. No. 3.4e-68;
Matches 297; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1509 ATGTTACTCAGAGAGGCAATGCTTTACTGAACTGGAAGAGCGGCTGACACACTGG 1568
DB 275 AGGTTACTCAGAGAGGCAATGCTTTACTGAACTGGAAGAGCGGCTGACACACTGG 334
QY 1569 TCCATATTTCTTTTCCCAAGATCTCGGGAGCTTCGGAGCACTGTGCTACATCATCAT 1628
DB 335 TCCATATTTCTTTTCCCAAGATCTCGGGAGCTTCGGAGCACTGTGCTACATCATCAT 394
QY 1529 CTAGGGATGCTCACCCTACTGCTGCTGCGCCACCTGAGCGGCTCCAGCCCTTCTGCT 1688
DB 395 CTAGGGATGCTCACCCTACTGCTGCTGCGCCACCTGAGCGGCTCCAGCCCTTCTGCT 454
QY 1689 GCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1748
DB 455 GCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 514
QY 1749 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1808
DB 515 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 574
QY 1809 CTACTT 1814
DB 575 CTACTT 580

RESULT 6
ID ABR51681 standard; DNA; 1920 BP.
XX ABR51681;
XX
XX 30-JUL-2002 (first entry)
XX
XX DNA encoding human ABCG5 protein.
XX
XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
XX arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
XX chromosome 2p21; ds.
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX CDS 1..1920
XX FT /*lag- a
XX FT /product- "Human ABCG5 protein"
XX FT /transl_except- (pos: 4..9, aa: GDSLSPFGSGMGL)
XX FT /note- "This sequence contains 13 exons"
XX
XX MO202227016-A2.

PD 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX
XX 25-SEP-2000; 2000US-235268P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (PATEL) PATEL S B.
XX (DEAN/) DEAN M.
XX Patel SB, Dean M;
XX WPI: 2002-416483/44.
XX P-PSDB: AAU98984.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
XX nucleic acid encoding the polypeptide, useful for treating
XX sitosterolemia, arteriosclerosis and heart diseases -
XX
XX Claim 38; Page 36-37; 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal, where an increase or decrease in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or decreases ABCG5 activity respectively. The cell culture or mammal comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The ABCG5 biological activity, or level of ABCG5 mRNA, or level of the polypeptide in a cell culture or mammal is also compared with that of a second cell culture or mammal comprising a wild type ABCG5 polypeptide. Stimulation of ABCG5 activity is useful for treating or preventing hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's disease. The method of the invention is useful for increasing cholesterol excretion and/or decreasing cholesterol adsorption. The present nucleic acid sequence represents the human ABCG5 gene located on chromosome 2p21. This sequence encodes the human ABCG5 protein of the invention.

Query Match 7.6%; Score 203.6; DB 24; Length 1920;
Best Local Similarity 54.4%; Pred. No. 3.4e-44;
Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

QY 335 GCTGCCAGAAATTTCTGTGAGCTGGGCAATCCAGAACTTCAAGTGAAGAGTGGC 394
DB 143 GCCGGAGAGAGTGGAGGAGGAGGAGTCTCAAAAGATGTCTCTTACGTGAGAGCGGGC 202
QY 395 AGATGCTGGCCATATATGAGGAGCTCAGGTTGTGGAGAGCCCTCTCTAGATGATCA 454
DB 203 AGATGCTGGCCATATATGAGGAGCTCAGGTTGTGGAGAGCCCTCTCTAGAGCCATGT 262
QY 455 CTGGCCGAGGTCACGCGGAGGAGTCAAGTCAAGTCAGGCGCATGTGATTAAGGGAGGCCA 514
DB 263 CCGGGAGAGCTGGGGCGGGGAGCTTCTGGGGAGGTATGTAAGCGCGGGCGC 322
QY 515 GCTGCGCTCAGCTGTGAGGAGAGTGTGCGCCAGCTGCGCCAGACACACAGCTGTCC 574
DB 323 TCGCGCGGAGAGCTTCCAGAGACTTCTCTACAGCTCGTCAGAGAGAGACACCTGTGA 382
QY 575 CCAACTGAGCTGCGGAGAGACCTTGGCTTCAATGCCAGATGGCGTCCAGAACTT 634
DB 383 GCAAGCTCAGCCGCGGAGACCTTGCATACACCGGCTGTGGCCATCGCGCGGGCA 442
QY 635 TCTGCCAGGCCACGCGGAGCAAAAGGTTGAGAGCGATCGGGAGAGCTGGCTTAAAGC 694
DB 443 ATCCCGGCTCTTCC---AGAGAGAGTGGAGGCGCTGATGAGAGCTGATGTAGCC 499

OY 695 AGTGGCTGACACCCCGCTGGCAACATGATGCGGGGTTGTGCGGGGTGAGCCGA 754
 DB 500 ATGTGGACGACCGACTGATGTGGCAACTGACCTTGGGGGATTTCCAGGGGTGAGCCG 559
 OY 755 GGAGAGTCAGATGGGGTGCAGCTCCTGTGTGAAACCCAGGAATCCTTATTCGACGAC 814
 DB 560 GCCGGGTCTCCATGCGACGCCAGCTGTCCAGAGATCCAGATGCTATGTTTATGTAGC 619
 OY 815 CCACCTCTGGGCTGACAGCTTTCACAGGCCCAACGCTGTAAGACCTTGTCCAGGCTGG 874
 DB 620 CAACACAGGCTGTGATGCTGATGCTGTAATGATGCTGCTGCTGCTGTAAGACTGG 679
 OY 875 CCAAGGCAACCGGCTGTGCTCAATCTCCCTGACAGCCCTGCTGCTGCTGCTGCTGCTG 934
 DB 680 CTCGACGAAACCGAATGTGTGTTCTCAACATTCACAGCCGCCCTTCTGAGCTTTTCAGC 739
 OY 935 TGTGTGATGCTGCTCTCTGATGACGCTGCTGACCCGCCATCTTAAAGGGGCGCCGAC 994
 DB 740 TCTTTGACAAATTTGCAATCTGCTGAGCTGCTGAGAGCTGATTTCTGTGGCAGCCAGCG 799
 OY 995 ACATGCTCAGATTTTCACAGCCATGCGGCTACCCCTGCTGCTGCTGCTGCTGCTGCTG 1054
 DB 800 AATGCTGATTTTTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 859
 OY 1055 ACTTCTATGCTGACCTGACAGCAATTCAGACGCGCAGCAGCAGCAGCAGCAGCAGCAGC 1114
 DB 860 ACTTCTATGCTGACCTGACAGCAATTCAGACGCGCAGCAGCAGCAGCAGCAGCAGCAGC 919
 OY 1115 GCGGAGAGCTGACG 1128
 DB 920 CCAAGAGAGCTCAG 933

RESULT 7

ID AAD22009 standard; DNA: 2340 BP.

AC AAD22009;

DT 12-FEB-2002 (first entry)

DE Human sitosterolemia susceptibility gene (SSG).

KM Human: sitosterolemia susceptibility gene; SSG: atherosclerosis;

KM sterol-related disorder; hyperlipidemia; hypercholesterolemia; therapy;

KM gall stone; coronary heart disease; cardiovascular disease; arthritis;

XX xanthoma; hemolytic anaemia; transgenic animal; chromosome 2p21; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 107..2062

FT Tag- a

XX /Product- "Human SSG protein"

XX MO200179272-A2.

XX 25-OCT-2001.

XX 18-APR-2001; 2001MO-US12758.

XX 18-APR-2000; 2000US-198465P.

XX 15-MAY-2000; 2000US-204234P.

XX (TULA-) TUTARIK INC.

XX Tian H, Schultz J, Shan B;

XX MPI: 2002-017598/02.

XX P-PSDB: AAE13290.

XX Novel sitosterolemia susceptibility gene polypeptide and

XX polynucleotide, useful for screening a compound that increases the

PT level of expression or activity of SSG polypeptide for treating

PT sterol-related disorder

PS Claim 8; Fig 8; 105pp; English.

The invention relates to an isolated sitosterolemia susceptibility gene (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP) binding cassette (ABC) family cholesterol transporter. SSG is useful for identifying a compound useful in the treatment or prevention of a sterol-related disorder, including sitosterolemia, hyperlipidemia, hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or nutritional deficiencies. SSG is also useful for treating cholesterol-associated diseases or conditions including coronary heart disease and other cardiovascular diseases, and sitosterolemia-associated condition including arthritis, xanthomas and chronic haemolytic anaemia. SSG expression cassette is useful in the production of transgenic non-human animals. SSG genes and their homologues are useful as tools for a number of applications including diagnosing sitosterolemia and other cardiovascular disorders, for forensics and paternity determinations, and for treating any of a large number of SSG associated diseases. The present sequence is human SSG DNA. Human SSG is located on chromosome 2p21.

Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other;

Query Match 7.64; Score 203.6; DB 24; Length 2340;

Best Local Similarity 54.44; Pred. No. 3.8e-44;

Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

OY 335 GCTGCGAATTTCTGTGATGCGGCAATCCAGAACCTTAAGTGAAGTGGCG 394
 DB 285 GCCGCGACGACGTGACAGCAGAGATCTCTCAAGATGCTCTTACGTAGAGCGCGC 344
 OY 395 AGATGCTGCCATCATAGGAGGAGCTCAGTGTGTGGAGAGCTCTCTGATGTGATCA 454
 DB 345 AGATCATGTGATCTTATGGAAGCTCAGAGCTCCGGGAACACAGCTCTGAGCCCATGT 404
 OY 455 CTGCGCGAGGTGACGCGGCGCAATCAATGAGCCAGATCTGATCAATGAGCGCA 514
 DB 405 CCGGAGAGCTGGGGCGCGCGGAGCTCTCTGAGGAGTGTATGTAAGCGCGCGCGC 464
 OY 515 GCTGCGCTCAGCTGTGAGAGAGTGTGGCCCGCCAGCCAGCAACAGCTGTCC 574
 DB 465 TCGCCCGGAGCACTTCCAGAGACTCTTCTCTACGCTCTGCAAGAGCAACCTGTGA 524
 OY 575 CCAACTTGTGCTGCGAGAGACTTGGCTTCAATGCGCAGATGCGGCTGCCAGACT 634
 DB 525 GAGCCCTACCGTTCGCGGAGAGCGTGCACATACCGCGCTGCGCATCCGCGCGCA 584
 OY 635 TCTCCAGGCCAGAGCTGACAAAGGTGAGAGAGTGTGCGGAGCTCGGCTTAAAGC 694
 DB 585 ATCCCGGCTCTCTCC---AGAGAAGAGTGAAGCGGCTCATGAGAGACTAGTGTAGCC 641
 OY 695 AGTGGCTGACACCCCGCTGGGCAACATGATGCTGCGGGGCTGTGCGGGGCTGAGCGCA 754
 DB 642 ATGTGCGACCGAGCTGATGATGCAATGAGCTTGGGGGCAATTTCCAGCTGTAGGCGC 701
 OY 755 GGAGAGTCAGATGGGGTGCAGCTCCTGTGGAACCCAGAGATCCTTATTCGAGCAAC 814
 DB 702 GCCGGTCTCCATGCGACGCCAGCTGTCTCAAGATCTTAAGCTCATGCTTTGTATGAGC 761
 OY 815 CCACCTCTGGGCTGACAGCTTCAACAGCCCAACCACTGTGAAGAGCTTGTCCAGGCTGG 874
 DB 762 CAACACAGGCTGTGATGCTGATGCTCTAATGAGATGTCGCTCTGCTGTGGAAGCTGG 821
 OY 875 CCAAGGCAACCGGCTGTGCTCATCTCTCTCAACAGCTGCTGCTGATCTTCAAGC 934
 DB 822 CTCGACGAAACCGAATTTGTGTCTACCAATTCACGAGCCCGCTGAGACTTTTCAGC 881
 OY 935 TGTGTGATGCTGCTCTCTGATGACGCTGCGACCCCGCATCTTAAAGGGGCGCGCCAGC 994
 DB 882 TCTTTGACAAATTTGCTATGCTTGTGGAAGCTGATTTTGTGCGACCGCGCGCG 941

XX OS Homo sapiens.
 XX XX WO200227016-A2.
 XX PD 04-APR-2002.
 XX PF 25-SEP-2001; 2001WO-US29859.
 XX PR 25-SEP-2000; 2000US-235268P.
 XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PA (PATEL S B.
 XX PA (DEAN/) DEAN M.
 XX PI Patel SB, Dean M.
 XX DR WPI: 2002-416483/44.
 XX PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT stitosterolemia, arteriosclerosis and heart diseases -
 XX XX Example 3; Page 37-38; 66pp; English.
 XX XX The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing stitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of human ABCG5 gene located
 CC on chromosome 2p21.
 CC XX
 S0 Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 other;
 Query Match 7.6%; Score 203.6; DB 24; Length 2516;
 Best Local Similarity 54.4%; Pred. No. 3.9e-44;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

QY 635 TCTCCAGGCCCCAGGCTGACAAAGGCTGGAGAGAGCTGATCCGGAGCTGCGCTTAGCC 694
 DB 619 ATCCCGGCTCTTCC--AGAAGAGGCTGGAGGCGTCAATGGCAGAGCTGACCTGAGCC 675
 QY 695 AGTGGCTGACACACCCGCGTGGGCAACATGATGAGCTGGGGGCTGGAGCGCA 754
 DB 676 ATGTGGCAGACCCGACTGATGGCAACTACACTTGGGGGCAATTTCCACAGGCTGAGCGC 735
 QY 755 GGAGAGTCAGACTGGGGTCACTCTGTGTGAACCCAGAAATCTTATTCGACGAAAC 814
 DB 736 GCCGGGTCTCATTCCAGAGCCAGCTGCTCCAGAGATCTTAAGTCAATGCTTGTGATGAGC 795
 QY 815 CCACCTGTGGGCTGCAGAGCTTACAGCCCAACCTGTGTGAGACCTTGTCCAGGCTGG 874
 DB 796 CAACCAAGGCTGACATGCTCAATGCTCAATGCTCTCTCTCTGCTGGAACCTGG 855
 QY 875 CCAAGGCAACCGGCTGGTCTCATCTCCCTCCACAGGCTGCTGACATCTCAGGC 934
 DB 856 CTCGAGAGAACCGAATTTGTTCTCACCATTACACAGCCCGTTCTGTGACTTTTCAGC 915
 QY 935 TGTGTGATCTGCTCTCTGATGAGCTGTGACCCCATCTACTTAAAGGCGGCGCCAGC 994
 DB 916 TCTTTGACAAATTTGCAATCTGAGCTTCCGAGACTGATTTTCTGTGGCAGCCAGCGG 975
 QY 995 ACATGCTCAGTATTTACAGCCATGCGCTACCCCTGTCTCTGCTACAGCAATCTGCTG 1054
 DB 976 AATCTCTGATTTCTTCAATGACTGCGTTACCTTGTCTGATCAATCAACCTTTTG 1035
 QY 1055 ACTTGTATGTCAGCTGACCGACATTTGACAGGCGCAGAGAGAGAGCAATTTGCCACCA 1114
 DB 1036 ACTTGTATGTCAGCTGACCGATGAGTGTGATPCCCAAGAGAGGGAATTAATTAACCT 1095
 QY 1115 GGGAGAGGCTCAG 1128
 DB 1096 CCAAGAGAGTCCAG 1109

RESULT 10
 ABR51686
 ID ABR51686 standard; cDNA; 2035 BP.
 XX
 AC ABR51686;
 XX
 AC 30-JUL-2002 (first entry)
 DT
 XX
 DE cDNA encoding rat ABCG5 protein.
 XX
 KW Rat; ABCG5; ATP-binding cassette gene 5; stitosterolemia; cholesterol; ss;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX
 OS Rat sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 8..1965
 FT /*tag= a
 FT /product= "Rat ABCG5 protein"

WO200227016-A2.
 04-APR-2002.
 25-SEP-2001; 2001WO-US29859.
 25-SEP-2000; 2000US-235268P.
 (USSH) US DEPT HEALTH & HUMAN SERVICES.
 (PATEL S B.
 (DEAN/) DEAN M.
 Patel SB, Dean M;
 WPI: 2002-416483/44.
 P-PSDB: AAU96986.

CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the mouse ABCG5 protein of the invention.

XX
 SO Sequence 1915 BP; 453 A; 502 C; 484 G; 476 T; 0 other:

Query Match 7.2%; Score 193.4; DB 24; Length 1915;
 Best Local Similarity 53.4%; Pred. No. 1.9e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

OY 335 GCTGCCAATTCCTGTGAGCTGGGCAATCCAGAACCTTAAGCTTCAAGTGAAGAGTGGGC 394
 DB 182 GCCAGCAGAGTGGGACAGGCAAAATCTCAAGATGCTCTCTTGTACATCGAGAGTGGCC 241
 OY 395 AGATGCTGGCCATCATAGGAGCTGAGGTGGGAGAGCCCTCTGCTAGATGATCA 454
 DB 242 AGATTATGTGATCTTAGGACAGCTCAGGCTCAGGAGACACACGCTGTGAGAGCCATCT 301
 OY 455 CTGGCCGAGTCAAGCGCGCAAGATCAAGTCAAGCCAGATCTGATCAATGAGGCCCA 514
 DB 302 CCGGAGGCTGCGGCGCACTGGAGCCCTGGAAGGGAGAGTGTGTTGTAATGGCTCCGAGC 361
 OY 515 GCTGCGCTCAGCTGTGTGAGAGTGTGTGGCCCACTGCGCAGCACAACAGCTGCTCC 574
 DB 362 TCGCGAGGAGCAGTCCAGACAGTCTCTCTCAGTCTCAGAGCAGCAGCTTTTCTGA 421
 OY 575 CCAACTTACTGTGAGAGAGCTTGGCCCTTATGCCCCAATGGGCTGCCAGAACT 634
 DB 422 GCAGCTCAGTGTGGCGAGAGCTTGGCATACAGC---GATGTGGCCCTTCTGCCCA 478
 OY 635 TCTCCAGCGCCAGCGTGAACAAAGGAGTGAAGAGTGAAGTGGGAGTGGGCTTAAAGC 694
 DB 479 GCTCCGCGGACTTCTACAAAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGC 538
 OY 695 AGTGGCTGACACCCGCTGGGCAATGATGATGCGGGGGTGTGCGGGGGTGAAGCCA 754
 DB 539 ACGTGGCGGCAAAATGATTTGGCAGTAAATTTTGGGGGAATTTCCAGTGGCAGCGGC 598
 OY 755 GGAGTCAAGATTTGGGGTGAAGTCTGAGTGAAGCCAGAAATCTTATTTTCAGCAAC 814
 DB 599 GCCAGTTTCCATGCAAGCCCACTCTTCAGAGACCCCAAGTCAATGATGATGATGAGC 658
 OY 815 CCACCTTGGGCTGAGAGCTTTCACAGCCCACTGAGTGAAGAGCTTGTCCAGGGCG 874
 DB 659 CAACACAGAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGC 718
 OY 875 CCAAGGCAACCGGCTGTGCTATCTCCACAGCCCTGCTGTGATCTTCAAGC 934
 DB 719 CTCGAGGAGCAATTTGATTTCTCAACATCCACAGCCCTGCTGAGCTTCTCAAC 778
 OY 935 TGTGATCTGCTCTCTGATGAGTGTGAGCCAGCCCACTTACTATTGAGGGGCGCCAGC 994
 DB 779 ACTTGAGCAAAATTCATCTCTGACTTACGAGAGTGTGTTGTGAGCCAGAGG 838
 OY 995 ACATGTCACATATTTCACAGCCATTCGCTACCCCTGCTGCTGCTGCTGCTGCTGCTG 1054
 DB 839 AGATGCTTGGCTTCTCAATTAATGCTGTTTACCCCTCTCTGCTGCTGCTGCTGCTGCT 898
 OY 1055 ACTTCTATGTGAGCTTCAAGCAATTCACAGCGCCAGAGAGAGAGAGAGAGATGGCCACA 1114
 DB 899 ATTATTACATGAGCTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 958
 OY 1115 GGGAGAGGCTCAGTCACTGCA 1137
 DB 959 ACAGAGAGTACAGATCTGGA 981

RESULT 12

AAD48880
 ID AAD48880 standard; DNA; 1959 BP.
 XX
 AC AAD48880;
 XX
 XX 24-MAR-2003 (first entry)
 DT
 XX
 DE Mouse ABCG5 DNA.
 XX
 XX ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolemia; hyperlipidemia; hypercholesterolemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW mouse; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH 1.1591
 FT CDS /tag=a
 FT /product="mABCG5 protein"

MO200281691-AZ.
 17-OCT-2002.
 20-NOV-2001; 2001WO-US43823.
 20-NOV-2000; 2000US-252235P.
 PR 28-NOV-2000; 2000US-253645P.
 XX
 XX (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI
 PI Hobbs HH, Shan B, Barnes R, Tlan H;
 DR WPI: 2003-058548/05.
 DR P-PSDB; AAE31702.
 XX
 XX New ABCG8 polypeptides and nucleic acids, useful for treating
 PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies -
 XX
 XX Claim 11; Page 73; 94pp; English.

XX The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is mouse
 CC ABCG5 DNA.

XX
 SO Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 other:

Query Match 7.2%; Score 193.4; DB 25; Length 1959;
 Best Local Similarity 53.4%; Pred. No. 2e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

OY 335 GCTGCCAATTCCTGTGAGCTGGGCAATCCAGAACCTTAAGCTTCAAGTGAAGAGTGGGC 394
 DB 182 GCCAGCAGAGTGGGACAGGCAAAATCTCAAGATGCTCTCTTGTACATCGAGAGTGGCC 241
 OY 395 AGATGCTGGCCATCATAGGAGCTGAGGTGGGAGAGCCCTCTGCTAGATGATCA 454
 DB 242 AGATTATGTGATCTTAGGACAGCTCAGGCTCAGGAGACACACGCTGTGAGAGCCATCT 301
 OY 455 CTGGCCGAGTCAAGCGCGCAAGATCAAGTCAAGCCAGATCTGATCAATGAGGCCCA 514
 DB 302 CCGGAGGCTGCGGCGCACTGGAGCCCTGGAAGGGAGAGTGTGTTGTAATGGCTCCGAGC 361

```

OY 515 GCTGCGCTCAGTGTGAGAGAGTGTGGCCGACGCGCCGACACACACAGCTGCTCC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 TGGCGAGGAGACAGTTCACAGACTGCTTCTCAGCTGCTGAGAGAGAGCTTTTCTGA 421
OY 575 CCAACTTACTGTGCGAGAGAGCTTGGCCCTTCAATGCGCGAGATGCGGCTCCAGAACCT 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 GCAGCTCAGTGTGCGGAGAGAGCTTGGATACAGC---GATGCTGCGCTTGTCCGCA 478
OY 635 TCTCCAGGCGCCAGCTGACAAAAGGTGAGAGAGCTGATGCGGAGCTGCGCTTAGCC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 GCTCGCGGAGCTTCTACACACAGAGAGAGAGCTGATGACAGAGCTGAGCTGAGCC 538
OY 695 AGTGGCTGACACCGCGTGGGAGACATGACTGCGGGGCTTGTGCGGGGCTGAGCGCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACCTGGCGGACCAATGATTTGGCAGCTATATATTTTGGGGAATTTCCAGTGGGAGCGC 598
OY 755 GGGAGTCAAGATTGGGCTGACAGCTCTGGAACCCAGAGATCTTATTTCTGAGGAC 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GCGAGATTTCATCGACGCGACACTCTTTCAGAGACCCCAAGCTCATGATGATGATAGC 658
OY 815 CCAGCTCTGGCTCGACAGCTTGCACAGCCCAACCTGTGAGAGACTTGTCCAGGCTGG 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 CAACCAAGAGAGCTGAGCTGATGACAAATCAATTTGCTTCTTGGCTGAGCTGG 718
OY 875 CCAAGGCAACCGGCGTGCATCTCCCTCCAGAGCTGCTGTGACATCTTACGC 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 CTGCGAGGAGCGAATGTGATTTGATTCACATCCACAGCTGCTGTGAGCTTCCACAC 778
OY 935 TGTGTGATCTGCTCTCTGATGACGTCGTGCAACCCCAATCACTAATGGGGGGCCAC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 ACTTGACAAATTCGATCTGATCTGACGAGAGTGTGCTGTGCGACCCCAAGG 838
OY 995 ACATGCTCAGATTTTACAGAGCCATCGCTACCCCTGCTGCTACAGCAATCTGCTG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 839 AGATCTGCTGCTTCTTCAATACGTGCTTACCCCTGCTGACATTTCCATCTTGG 898
OY 1055 ACTTGTATGAGCTGACCCACATTTGACAGCGCAGAGAGAGCAATGGCCACCA 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 ATTTTACATGAGCTGATGATGAGACACCAAGCAGAGAGCGGGAATAGAAACGT 958
OY 1115 GGGAGAGGCTCAGTCACTCGCA 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 959 ACAAGCGAGTACAGATGCTGGA 981

```

RESULT 13

AAd22008 standard; DNA; 2258 BP.

AAd22008;

12-FEB-2002 (first entry)

Mouse sitosterolemia susceptibility gene (SSG).

Mouse: sitosterolemia susceptibility gene; SSG; atherosclerosis;

steroid-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;

gall stone; coronary heart disease; cardiovascular disease; arthritis;

xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

Mus sp.

Key Location/Qualifiers

CDS 47..2005

FT /tag- a

PF 18-APR-2001; 2001WO-US12758.

```

XX 18-APR-2001; 2000US-198465P.
PR 15-MAY-2000; 2000US-204234P.
XX
PA (TULA-) TULARIK INC.
PI Tian H, Schultz J, Shan B;
DR MPI; 2002-017598/02.
DR P-PSDB; AAE13289.
XX
PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT steroid-related disorder
XX
PS Claim 8; Fig 7; 105bp; English.
XX
CC The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC steroid-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17.
XX
SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 other:

```

Query Match 7.2% Score 193.4; DB 24; Length 2258;

Best Local Similarity 53.4%; Pred. No. 21e-41;

Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

```

OY 335 GCTGCCAGAAATTTGTGAGTGGAGTGGACATCCAGAACCTTAAGCTCAAAAGTGAAGTGGG 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 GCCACAGAGAGTGGAGACAGCGCAAAATCTCAAAAGATGCTCTTGTACATCCAGAGTGGC 287
OY 395 AGATGCTGGCCATCATATAGGAGCTGAGTGTGGAGAGCCTCTTGTAGATGATCA 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 AGATTATGTGATCTTATAGCAGCTCAGGCTCAGGGAAGACACGCTGCTGGACGCCATCT 347
OY 455 CTGGCGAGGTACAGCGCGCAAGATCAAGTCAAGGCGCAATCTGATCAATGAGGACCCA 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 CCGGAGAGCTCGGCGCAGCTGGACCTGGAAGGAGGTGTGTGTGATGCTGCGAGC 407
OY 515 GCTGCGCTCAGCTGTGAGAGAGTGTGCGCCAGTGCAGCAGACACAGAGCTGCTCC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 TGGCAGAGGACAGTTCACAGACAGTCTTCTTCAAGTCTTCAAGACGACCTTTTCTGA 467
OY 575 CCAACTTACTGTGCGAGAGAGCTTGGCCCTTCAATGCCAGATGCGGCTGCCAGAACCT 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 GCAGCTTCACTGTGCGGAGAGCTTGGATACACAGC---GATGCTGGCCCTTGGCGCA 524
OY 635 TCTCCAGGCGCCAGCTGACAAAAGGGTGGAGAGAGCTGATGCGGAGCTGCGCTTAGCC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 GCTCGGCGGACTTTACAAACAGAGATGAGGAGATGACAGAGCTGAGCCTTAGGCC 584
OY 695 AGTGGCTGACACCGCGGTGGGCAACATGTACGTGCGGGGCTTGTGCGGGGTGAGCGCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 ACGTGGCGGACCAATGATTTGAGCAGCTATATTTTGGGGAATTTCCAGTGGCAGCGCC 644
OY 755 GGAGAGTCAAGATTTGGGCTGAGCTCTGTGGAACCCAGGAATCTTATTTCTGAGCAAC 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 GCGAGTTTTCATGCGAGCCCACTCTTCAAGAGACCCCAAGATCATGATGATGAGC 704

```

OY 815 CCACCTGCGGCTCGACAGCTTCACAGCCACACACCTGCTGTAAGACCTTGTCCAGCTGG 874
 DB 705 CAACACAGAGACTGACATGACATGACAAATATGCTCTTGTGGCTGAGCTGG 764
 OY 875 CCAAGGACACCGGCTGCTGCTCATCTCCCTCCACAGCCCTGCTGACATCTTCCAGGC 934
 DB 765 CTCGAGGAGCCGAATGTGATTTGATTCACCATCCACAGCCCTGCTGAGCTCTTCCAC 824
 OY 935 TGTGTATCTGCTCTCCGATGAGCTGCGACCCCATCTGCTAGAGGCGGCCAGC 994
 DB 825 ACTTCGACAAATTTCCATCTCTGACTTACGAGAGAGTGTGCTGTGCGACCCAGAGG 884
 OY 995 ACATGCTCAGATTTTTCACAGCCATCGGCTACCCCTGCTGCTGCTACAGCATCTGCTG 1054
 DB 885 AGATGCTGCTGCTCTTCAATACGTGTGTTACCCCTGCTGCTGACATCTTCCATCTGCTT 944
 OY 1055 ACTTCTATGTGACCTGACACGACATTTGACAGCGGACAGAGAGCAATTTGGCCACCA 1114
 DB 945 ATTTTACATGAGCTTGTGACATCAGTGACACCCAAAGACAGAGCGGGAATAGAAAGCT 1004
 OY 1115 GGGAGAAAGCTCAGTCACTGCA 1137
 DB 1005 ACAAGCGAGTACAGATCTGGAA 1027
 RESULT 14
 ABR51685
 ID ABR51685 standard; cDNA: 2354 BP.
 AC ABR51685;
 XX 30-JUL-2002 (first entry)
 DE Mouse ABCG5 cDNA: sequence.
 XX Mouse; ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM ss.
 XX Mus sp.
 OS W0200227016-A2.
 PN 04-APR-2002.
 PD 25-SEP-2001; 2001MO-US29859.
 PE 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL) PATEL S. B.
 PA (DEAN) DEAN M.
 PI Patel SB, Dean M;
 PI WPI: 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 XX nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS Example 3; Page 45; 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or

CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of
 CC the invention.
 XX
 SO Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 other;
 Query Match 7.2%; Score 193.4; DB 24; Length 2354;
 Best Local Similarity 53.4%; Pred. No. 2.2e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;
 OY 335 GCTGCCAATTTCTGTAGCTGGGATCCAGAACTTAACTTCAAGTGAAGAGTGGC 394
 DB 320 GCCAGCAGAACTGGGACAGGCAAAATCTCAAGATGTCTCTTGTACATCGAGAGTGGC 379
 OY 395 AGATGCTGGCATCATAGGAGCTTCAGGTTGGGAGAGCCCTTGTAGTGTGATCA 454
 DB 380 AGATTATGTGATCTTAGGACGCTCAGGCTCAGGAGACACAGCTGCTGGACGCTATCT 439
 OY 455 CTGGCCGAGGCTGACGGCGCAATCAAGTCAAGTCAAGGCTGATGATGAGGAGCCCA 514
 DB 440 CCGGAGAGCTGCGGGCAGCTGGAGACCTGGAAGGAGAGTGTGATGATGCTGGAGC 499
 OY 515 GCTGGCCCTCAGCTGTGAGGAAGTGTGCGCCACGTGCGCAGACACAGCTGCTCC 574
 DB 500 TCGCAGAGGACCAATTCACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 559
 OY 575 CCAACTGACTGTGAGAGAGCTTGGGCTTCAATTTGCCACATGCGGCTGGCCAGAACT 634
 DB 560 GCACCTCAGCTGTGAGAGAGCTTGGGCTTCAATTTGCCACATGCGGCTGGCCAGAACT 616
 OY 635 TCTCCAGAGCCCAAGCTGACAAAGAGGTGAGAGAGCTGATGCGGAGCTGCGCTTAA 694
 DB 617 GCTCGGGAGCTTGTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 676
 OY 695 AGTGGCTGACACCCGCTGGGCAATGATGATGATGATGATGATGATGATGATGATGAT 754
 DB 677 AGTGGCTGACACCCGCTGGGCAATGATGATGATGATGATGATGATGATGATGATGAT 736
 OY 755 GGAGAGTCAAGATGGGGTGGAGCTGCTGCTGAGAGCCAGAAATCTTATTTCTGACAGAC 814
 DB 737 GCCAGATTTCCATGCGAGCCCACTCTGAGAGCCCAAGATGATGATGATGATGATGATG 796
 OY 815 CCACCTTGGGCTGACAGCTTTCACAGCCCAACTGCTGTAAGACCTTGTCCAGCTGG 874
 DB 797 CAACACAGAGACTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 856
 OY 875 CCAAGGACACCGGCTGCTGCTCATCTCCCTCCACAGCCCTGCTGACATCTTCCAGGC 934
 DB 857 CTCGAGGAGCCGAATGTGATTTGATTCACATCCACAGCCCTGCTGAGCTCTTCCAC 916
 OY 935 TGTGTATCTGCTCTCCGATGAGCTGCGACCCCATCTGCTAGAGGCGGCCAGC 994
 DB 917 ACTTCGACAAATTTCCATCTCTGACTTACGAGAGAGTGTGCTGTGCGACCCAGAGG 976
 OY 995 ACATGCTCAGATTTTTCACAGCCATCGGCTACCCCTGCTGCTGCTGCTGCTGCTGCTG 1054
 DB 977 AGATGCTGCTGCTCTTCAATACGTGTGTTACCCCTGCTGCTGCTGCTGCTGCTGCTG 1036
 OY 1055 ACTTCTATGTGACCTGACACGATTTGACAGGGGACAGAGAGCAATTTGGCCACCA 1114
 DB 1037 ATTTTACATGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1096
 OY 1115 GGGAGAAAGCTCAGTCACTGCA 1137
 DB 1097 ACAAGCGAGTACAGATCTGGAA 1119

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)
11512.153 Million cell updates/sec

sequence: 1 gtgtccctgtctccaggaac.....caataaatgtattgagc 2669

Searched: 2888711 seqs, 20454813386 residues

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

GenEmb1:*
1: gb_ba:*

2	go_hng:
3	gb_in:
4	gb_om:
5	gb_ov:
6	gb_pat:
7	gb_ph:
8	gb_pl:
9	gb_pr:
10	gb_ro:
11	gb_sts:
12	gb_sy:
13	gb_un:
14	gb_vl:
15	em_ba:
16	em_fun:
17	em_hun:
18	em_in:
19	em_mu:
20	em_om:
21	em_or:
22	em_ov:
23	em_pat:
24	em_pl:
25	em_pl:
26	em_ro:
27	em_sts:
28	em_un:
29	em_vl:
30	em_hng_hum:
31	em_hng_inv:
32	em_hng_other:
33	em_hng_mus:
34	em_hng_pln:
35	em_hng_rod:
36	em_hng_mem:
37	em_hng_vrt:
38	em_sy:
39	em_hngo_hum:
40	em_hngo_mus:
41	em_hngo_other:

No.	Result	Score	Query Match	Length	DB	ID	Description
1	2669	100.0	2669	6	AX685735	AX685735 Sequence	
2	2655.2	99.5	2679	9	AF324494	AF324494 Homo sapi	
3	2020.4	75.7	2022	9	AF320294	AF320294 Homo sapi	
4	1680.6	63.0	3239	6	AX478099	AX478099 Sequence	
5	1518.6	56.9	4829	10	AF351785	AF351785 Ratlus no	
6	1511.6	56.6	3674	10	AX324495	AF324495 Mus muscu	
7	1454.2	54.5	2284	10	AY196216	AY196216 Mus muscu	
8	1449.4	53.6	2285	10	AX196215	AX196215 Mus muscu	
9	1430	53.6	2019	6	AX685731	AX685731 Sequence	
10	724	27.1	127066	9	AC084265	AC084265 Homo sapi	
11	724	27.1	182261	2	AC087053	AC087053 Homo sapi	
12	722.4	27.1	139342	9	AC108476	AC108476 Homo sapi	
13	663.6	24.9	2201	9	AF35182813	AF35182813 Homo sapi	
14	274.2	10.3	660	9	AF351812S06	AF351812 Homo sapi	
15	273.4	10.2	1292	9	AF351812S11	AF351812 Homo sapi	
C	260.6	9.8	68166	2	AC084712	AC084712 Homo sapi	
16	244	9.1	1323	9	F351812S04	F3518125 Homo sapi	
17	231.2	8.7	250	11	G329195	G329195 human STS S	
C	230.4	8.6	237445	2	AC120701	AC120701 Ratlus no	
19	230.4	8.6	312858	2	AC112747	AC112747 Ratlus no	
C	229.4	8.6	40929	10	AY145899	AY145899 Ratlus no	
21	226.8	8.5	1387	10	AF35199S06	AF351804 Mus muscu	
22	207.4	7.8	156829	2	AC120532	AC120532 Oryza sat	
C	204.8	7.7	1378	10	F351799S11	F351809 Mus muscu	
24	204.8	7.7	185545	2	AC122243	AC122243 Mus muscu	
C	204.8	7.6	685	9	F351812S09	AF351820 Homo sapi	
26	203.8	7.6	1920	6	AX456519	AX456519 Sequence	
27	203.6	7.6	2340	6	AX685733	AX320883 Sequence	
28	203.6	7.6	2340	6	AF320293	AX685733 Sequence	
29	203.6	7.6	2340	6	AF320293	AF320293 Homo sapi	
30	203.6	7.6	2516	6	AX456520	AX456520 Sequence	
31	203.6	7.6	2740	9	AF312715	AF312715 Homo sapi	
32	203.6	7.6	400	11	G16507	G16507 human STS S	
C	202.4	7.6	2351	10	AY195873	AY195873 Mus muscu	
34	195	7.3	2035	6	AX456526	AX456526 Sequence	
35	194.4	7.3	2470	10	AF312714	AF312714 Ratlus no	
36	194.4	7.3	2351	10	AY195872	AY195872 Mus muscu	
37	194	7.2	1915	6	AX456523	AX456523 Sequence	
38	193.4	7.2	1915	6	AX685729	AX685729 Sequence	
39	193.4	7.2	2258	6	AX320881	AX320881 Sequence	
40	193.4	7.2	2354	6	AX456524	AX456524 Sequence	
41	193.4	7.2	2354	10	AF312713	AF312713 Mus muscu	
42	193.4	7.1	1716	10	F351799S13	AF351811 Mus muscu	
43	190.4	6.9	1470	10	F351799S04	AF351802 Mus muscu	
44	184.2	6.5	1069	6	AX456528	AX456528 Sequence	
45	173.2						

AX685735

LOCUS	AX685735	2669 bp
DEFINITION	Sequence 7 from Patent WO02081691	
ACCESSION	AY055735	

KEYWORD	AX685735	GI:29371744
VERSION	AX685735.1	
ACCESSION	AX685735	

SOURCE ORGANISM	Homo sapiens (human) Homo sapiens
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

REFERENCE

AUTHORS Hobbs, H. H., Shan, B., Barnes, R. and Tian, H.
TITLE Abcg5 and abcg8: compositions and methods of use
JOURNAL Patent: WO 02081691-A 7 17-OCT-2002;

Tularik Inc. (us) ; BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(us)

FEATURES
Location/Qualifiers
1. 2669

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
100..2121

CDS

/note="unnamed protein product; human ABCG8 (hABCG8)"
/codon_start=1
/protein_id="CAD86573.1"
/db_xref="GI:29371745"

/translation="MAGKAERGLPKATPDITSGLDRLFSSRSNLSLFTYSGOP
NLEVRDLNIOVDLSQVPEEQIAQFMPTSPSCONSCEIGIONLSFKVASQMLA
IIGSGGSRASLADYITGRHGKIKSQIWMNGPSSPOLVRKCAVAROHMLLP
LVRETLAIFNOMRLPTFSQARDKRYEDVLAELRLKQADRTGVNMYRGLSGER
RRVSGVOLNMRGILIDEPISGLDSTFANHLVKTSLRKNLNLVLSIHQPSDI
PLFLVLNLTSGPFIYGAOHMVOYETALGRCPRSNADPVDLTSDRREBO
PLATREKQSLAFLLEKRYRDLDFLMAKTRDLDEDECVESVTPPLDNLCPSTKM
PEAVQFTLLRQISNDPRDLPLILHGAENCLSMITGFLYRFGHSIOLSPMDTA
LLEMLGALIPFNVLIDVLSKCYSERAMLYELEDGLYTTGPEFAKILGELPEHCAYI
IYGMPTVLANLRGLQPLHLHLVWLIVVCCRIALAAALAPTEPHMASFENAL
YNSFLAGCFMNLSSLTWTPANISKVSFLMCFGLMKIOPSRRTYKMLGNLITAV
SGDKILSMELDSYPLVAILYIVIGLSGFNVLVYSLRPTKORSQW"

BASE COUNT 595 a 768 c 722 g 584 t

Query Match 100.0%; Score 2669; DB 6; Length 2669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGTCCCTGCTCCAGAAACAGATGAGACACTGGCCCTGGCAGCAGCAGCTGGCT 60
DB 1 GTGTCCCTGCTCCAGAAACAGATGAGACACTGGCCCTGGCAGCAGCAGCTGGCT 60
OY 61 AAGAGAGCTGAGCCAGGCTCAGACCTGTGGGCCCCCATGGCCGGAAGGCGGCAAG 120
DB 61 AAGAGAGCTGAGCCAGGCTCAGACCTGTGGGCCCCCATGGCCGGAAGGCGGCAAG 120
OY 121 GAGAGAGGCTGCGCAAGAGGGGCCACTCCAGATACCTCGGCTCCAGAGATGATG 180
DB 121 GAGAGAGGCTGCGCAAGAGGGGCCACTCCAGATACCTCGGCTCCAGAGATGATG 180
OY 181 TTTCCTCTGAAAGTGAACAGACCTGTACTTCACTTCACTTCACTTCACTTCACTT 240
DB 181 TTTCCTCTGAAAGTGAACAGACCTGTACTTCACTTCACTTCACTTCACTTCACTT 240
OY 241 GAGTCAGAGACCTTAACTACAGAGTGAAGTGGCCCTCTCAGATGCTTGGTTGAGCAG 300
DB 241 GAGTCAGAGACCTTAACTACAGAGTGAAGTGGCCCTCTCAGATGCTTGGTTGAGCAG 300
OY 301 CTGGCTCAGTTCAGAGATCCCTGGACATCTCCAGAGCTCCAGAAATTTCTTGGAGCTGGGC 360
DB 301 CTGGCTCAGTTCAGAGATCCCTGGACATCTCCAGAGCTCCAGAAATTTCTTGGAGCTGGGC 360
OY 361 ATTCAGAACCTTAACTTAAAGTGAAGTGGGCAAGATGCGCATCATATAGGAGAGCTCA 420
DB 361 ATTCAGAACCTTAACTTAAAGTGAAGTGGGCAAGATGCGCATCATATAGGAGAGCTCA 420
OY 421 GGTGTGGAGAGGCTCTTGTCTAGATGTGATCACTGGCCGAGGTCACGGGCGCAAGATC 480
DB 421 GGTGTGGAGAGGCTCTTGTCTAGATGTGATCACTGGCCGAGGTCACGGGCGCAAGATC 480
OY 481 AAGTCAGGCGAGATCTGATCAATGGGAGCCAGAGCTGCTCAGCTGTGATGAGAGTGT 540
DB 481 AAGTCAGGCGAGATCTGATCAATGGGAGCCAGAGCTGCTCAGCTGTGATGAGAGTGT 540
OY 541 GTGGCCCAAGTGGCCGAGACACAGAGTGTCTCCCAACTTGAATGTGGCAGAGACCTTG 600
DB 541 GTGGCCCAAGTGGCCGAGACACAGAGTGTCTCCCAACTTGAATGTGGCAGAGACCTTG 600
OY 601 GCCTTCATTTGCCAGATGCGGCTGCCAGAACTTTCTCCAGGCGGAGCAAAAGG 660
DB 601 GCCTTCATTTGCCAGATGCGGCTGCCAGAACTTTCTCCAGGCGGAGCAAAAGG 660

DB 601 GCCTTCATTTGCCAGATGCGGCTGCCAGAACTTTCTCCAGGCGGAGCAAAAGG 660
OY 661 GTGAGAGAGTATGCGGAGAGCTGGCGGCTTAGAGAGTGGAGACACCGGCTGGGCAAC 720
DB 661 GTGAGAGAGTATGCGGAGAGCTGGCGGCTTAGAGAGTGGAGACACCGGCTGGGCAAC 720
OY 721 ATGTACGTGGGGGGTGTGTGGGGGGGTGAGCCAGAGAGTACAGATTGGGTGGCAGCTC 780
DB 721 ATGTACGTGGGGGGTGTGTGGGGGGGTGAGCCAGAGAGTACAGATTGGGTGGCAGCTC 780
OY 781 CTGTGGAAACCCAGGAATCTTATTTCTGACAGACACCACTCTGGGCTGACAGCTTACCA 840
DB 781 CTGTGGAAACCCAGGAATCTTATTTCTGACAGACACCACTCTGGGCTGACAGCTTACCA 840
OY 841 GCCCAACACCTGGTGAAGACCTTTCAGAGGCTGGCCAAAGCAACCGGCTGGTGTCTATC 900
DB 841 GCCCAACACCTGGTGAAGACCTTTCAGAGGCTGGCCAAAGCAACCGGCTGGTGTCTATC 900
OY 901 TCCCTCCACAGACCTGCTGTGACATCTTTCAGAGGCTGTTGATCTGCTCTGATGACG 960
DB 901 TCCCTCCACAGACCTGCTGTGACATCTTTCAGAGGCTGTTGATCTGCTCTGATGACG 960
OY 961 TCTGGCACCCCATCTACTTATGAGGGGCGGCCAGACATGCTCAGATATTTCACAGCCATC 1020
DB 961 TCTGGCACCCCATCTACTTATGAGGGGCGGCCAGACATGCTCAGATATTTCACAGCCATC 1020
OY 1021 GGCCTACCCCTGCTGCTGCTACAGCAATGCTGCTCTATGAGGAGCTTACCAAGCAT 1080
DB 1021 GGCCTACCCCTGCTGCTGCTACAGCAATGCTGCTCTATGAGGAGCTTACCAAGCAT 1080
OY 1081 GACAGGCGCAGACAGAGCAGGAATTTGGCCACAGAGGAGAGGCTCACTCACTGCAGACC 1140
DB 1081 GACAGGCGCAGACAGAGCAGGAATTTGGCCACAGAGGAGAGGCTCACTCACTGCAGACC 1140
OY 1141 CTGTTTCTAGAAAAGTCCGTGACTTATGATGATCTTATGAAAAGCAGAGCAAGAT 1200
DB 1141 CTGTTTCTAGAAAAGTCCGTGACTTATGATGATCTTATGAAAAGCAGAGCAAGAT 1200
OY 1201 CTGAGCAGAGACCTGTGTGAAAAGCAGGCGGACCCACTAGACACCACTGGCTCCCG 1260
DB 1201 CTGAGCAGAGACCTGTGTGAAAAGCAGGCGGACCCACTAGACACCACTGGCTCCCG 1260
OY 1261 AGTCTTACGAGAGATGCTGAGGAGGCTGAGCAGAGTTCAGAGCTGATCCGTCAGATG 1320
DB 1261 AGTCTTACGAGAGATGCTGAGGAGGCTGAGCAGAGTTCAGAGCTGATCCGTCAGATG 1320
OY 1321 TCCAAAGACCTCCAGAGACCTGCCACCTCTCATCCATGAGGGGAGGCTGTCTGATG 1380
DB 1321 TCCAAAGACCTCCAGAGACCTGCCACCTCTCATCCATGAGGGGAGGCTGTCTGATG 1380
OY 1381 TCAATGACATGCGCTCTCTATTTTGGCATGAGGAGCATTCAGCTCTCTCATGAGAT 1440
DB 1381 TCAATGACATGCGCTCTCTATTTTGGCATGAGGAGCATTCAGCTCTCTCATGAGAT 1440
OY 1441 ACAGCCGCGCTTGTGTATGATGCTGCTCATCCCTTCAATGAGGGGAGGCTGTCTGATG 1500
DB 1441 ACAGCCGCGCTTGTGTATGATGCTGCTCATCCCTTCAATGAGGGGAGGCTGTCTGATG 1500
OY 1501 ATTCGCAAAATGTTACTGAGAGAGGCAATGCTTACTATGTAAGTGAAGAGGCTGTAC 1560
DB 1501 ATTCGCAAAATGTTACTGAGAGAGGCAATGCTTACTATGTAAGTGAAGAGGCTGTAC 1560
OY 1561 ACCACTGGTCCATATTTCTTGGCAAGATCCCTGGGGAGCTTGGGAGCACTGTGGCTAC 1620
DB 1561 ACCACTGGTCCATATTTCTTGGCAAGATCCCTGGGGAGCTTGGGAGCACTGTGGCTAC 1620
OY 1621 ATCATTCATTAAGGAGATGCCACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 ATCATTCATTAAGGAGATGCCACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
OY 1681 TTCTGCTGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
DB 1681 TTCTGCTGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740

QY 1741 GCCGCCGCCGCTCTCCACCTTCACATGCGCTCTCTTCAGCAATGCCCTCTAC 1800
 |||||
 DB 1741 GCCGCCGCCGCTCTCTCCACCTTCACATGCGCTCTCTTCAGCAATGCCCTCTAC 1800
 |||||
 QY 1801 AACTCTCTTACCTGCGCGGGGCTTCATGATTAACCTTGACAGCCCTGTGACAGTCC 1860
 |||||
 DB 1801 AACTCTCTTACCTGCGCGGGGCTTCATGATTAACCTTGACAGCCCTGTGACAGTCC 1860
 |||||
 QY 1861 GCGTGAATTTCCAAAGTGTCTCTCGGGGGTGTGTTGAGGGCTGTGTAAGTTCAG 1920
 |||||
 DB 1861 GCGTGAATTTCCAAAGTGTCTCTCGGGGGTGTGTTGAGGGCTGTGTAAGTTCAG 1920
 |||||
 QY 1921 TTCACAGAGAAGAACTTATAAATGCTCTCGGGAACTCACCATCGCGTCTCAGAGAT 1980
 |||||
 DB 1921 TTCACAGAGAAGAACTTATAAATGCTCTCGGGAACTCACCATCGCGTCTCAGAGAT 1980
 |||||
 QY 1981 AAAATCTCAGTGCATGAGAGCTGAGCTGACCTCTCTACGCCATCTACCTCATCTC 2040
 |||||
 DB 1981 AAAATCTCAGTGCATGAGAGCTGAGCTGACCTCTCTACGCCATCTACCTCATCTC 2040
 |||||
 QY 2041 ATTGGCTCAGGGGCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2100
 |||||
 DB 2041 ATTGGCTCAGGGGCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2100
 |||||
 QY 2101 AAACCAAGTCAAGAGCTGATTCACGCCAGACGCTCTGCGCGCTGTGGGGAGCTGAGC 2160
 |||||
 DB 2101 AAACCAAGTCAAGAGCTGATTCACGCCAGACGCTCTGCGCGCTGTGGGGAGCTGAGC 2160
 |||||
 QY 2161 AGACCTTCAACTGACCT 2220
 |||||
 DB 2161 AGACCTTCAACTGACCT 2220
 |||||
 QY 2221 CCTTACAGATGCTGCTACATCCGCGCCAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCT 2280
 |||||
 DB 2221 CCTTACAGATGCTGCTACATCCGCGCCAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCT 2280
 |||||
 QY 2281 GATGCACTAGAAATTAAGACAGTGCAGAAAGGATTTCTGCTCACTGGAGAGCTGCAT 2340
 |||||
 DB 2281 GATGCACTAGAAATTAAGACAGTGCAGAAAGGATTTCTGCTCACTGGAGAGCTGCAT 2340
 |||||
 QY 2341 GACTGGGAGAAACCTGCACTCGGTGGGCACTACAACTGCTTAATTTCTCTTTGA 2400
 |||||
 DB 2341 GACTGGGAGAAACCTGCACTCGGTGGGCACTACAACTGCTTAATTTCTCTTTGA 2400
 |||||
 QY 2401 TATCATTTATATAGGCACTCGATATGATGGAGCAAACTAGGAATGGAATGGTAG 2460
 |||||
 DB 2401 TATCATTTATATAGGCACTCGATATGATGGAGCAAACTAGGAATGGAATGGTAG 2460
 |||||
 QY 2461 CTAGACTGTGAGAGATTTGGAACCTGAGGGAACAACTAGTACTGACGATTTG 2520
 |||||
 DB 2461 CTAGACTGTGAGAGATTTGGAACCTGAGGGAACAACTAGTACTGACGATTTG 2520
 |||||
 QY 2521 GCTTCATCTTTCAGAGGGGCCACACTCGGTGGTGGAGCCATCAATACAGAAAGTAC 2580
 |||||
 DB 2521 GCTTCATCTTTCAGAGGGGCCACACTCGGTGGTGGAGCCATCAATACAGAAAGTAC 2580
 |||||
 QY 2581 TAAATGTACAGAGAAAGTGCATCTCTCTTTTGTGTGGGTGATGGGCTCCAAAAGC 2640
 |||||
 DB 2581 TAAATGTACAGAGAAAGTGCATCTCTCTTTTGTGTGGGTGATGGGCTCCAAAAGC 2640
 |||||
 QY 2641 CAACGTGAACAATTAATAATGTATTGAGC 2669
 |||||
 DB 2641 CAACGTGAACAATTAATAATGTATTGAGC 2669
 |||||

RESULT 2
 AF324494
 LOCUS Homo sapiens sterol-in-2 (ABCG8) mRNA, complete cds.
 DEFINITION
 ACCESSION AF324494.1 GI:15088539
 VERSION
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS
 Lu, K., Lee, M.-H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
 Ose, L., Stalenhoef, A.F., Mettinen, T., Bjorkhem, I., Bruckert, E.,
 Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Sirtavasa, A. and
 Patel, S.B.
 TITLE
 Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterol-in-1 and
 sterol-in-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL
 MEDLINE
 PUBMED
 11452359
 2 (bases 1 to 2679)
 REFERENCE
 AUTHORS
 Lu, K., Lee, M.-H. and Patel, S.B.
 TITLE
 Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 Street, STB541, Charleston, SC 29403, USA
 JOURNAL
 FEATURES
 source
 1..2679
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2p21: between D2S2294 and D2S2298"
 /tissue_type="liver"
 1..2679
 /gene="ABCG8"
 91..2112
 /gene="ABCG8"
 /codon_start=1
 /product="sterol-in-2"
 /protein_id="AAK84078.1"
 /db_xref="GI:15088540"
 /translation="MAGKAERELGPKATPDPTSGLODRLEFSSEDSNLYFTYSGOP
 NLEVRDNCVDLADVPMELOAEFMPMTSPSCONCELGIONLSFKYSGOMLA
 IIGSGGCRALILVTGREGHGRKISQIVINQPSPOVRKCVAVRHONOLLPN
 LVRETLAFLIOMRLPRTFSOARDKREVDIARLRQCADTVGNNYVRGLSGGER
 RRSISGVILNPNPILIDPRTSGDSTAHNLVKTLSRLKRNRLVLSLQPRSDI
 FRPDIIVLHNSGPPIYLGAOHVQFTTAIGTFCPSINADTVLDSIDRSREQ
 ELATREKQSLAALFLEKVRDLDPFLHGAECALMSMTIGFLYGHGSIQSFPTKA
 PGAVQOFTLLIRROISNDFRDLPLHGAECALMSMTIGFLYGHGSIQSFPTKA
 LLFMIGALIPFNVLIDVSKYSERAMLYELEGLVTTGYPFAKILGELEPHQAVI
 ILYGPTWYLANLRPGLFHLFLVAVVPCGRIMALAAALILPFPMAFSPSNAL
 YNSFLYAGGEMINISLMTVPAMYSKYSFLMPCREGELMKIOPSRRTYKMPKGNLTIAV
 SGRKILSMELDSIPVLAITYIVIGLSGGFVNLVYSLEFTKPKPSDWM"
 BASE COUNT 609 a 767 c 720 g 583 t
 ORIGIN
 Query Match 99.5%; Score 2655.2; DB 9; Length 2679;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2657; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 10 CTCGAGAAACAGAGTGAAGACTGCGCTCGGAGCAGCAGCTGGTCTTAAGAGACT 69
 |||||
 DB 1 CTCGAGAAACAGAGTGAAGACTGCGCTCGGAGCAGCAGCTGGTCTTAAGAGACT 60
 |||||
 QY 70 GCACCCGAGGCTCAGACCTGTGTGGCCCATGTGCGCGGGAAGCGCGCAGAGAGAGAGG 129
 |||||
 DB 61 GCACCCGAGGCTCAGACCTGTGTGGCCCATGTGCGCGGGAAGCGCGCAGAGAGAGAGG 120
 |||||
 QY 130 CTGCCGAAAGGGGCGCAGTCCCGCAGATACCTCGGGGCTCCAGATATTTCTCTCTCT 189
 |||||
 DB 121 CTGCCGAAAGGGGCGCAGTCCCGCAGATACCTCGGGGCTCCAGATATTTCTCTCTCT 180
 |||||
 QY 190 GAAAGTGAACAACACCTGTACTTACCTACAGTGGGCCAGCCACACCTGGAGGTGAGA 249
 |||||
 DB 181 GAAAGTGAACAACACCTGTACTTACCTACAGTGGGCCAGCCACACCTGGAGGTGAGA 240
 |||||
 QY 250 GACCTCAACTACAGGTGAGACTGCGCTCTCAGAGTCCCTGTTGAGCAGACCTGCTCAG 309
 |||||

Db 241 GACCTCACTGCGAGTGGACCTGGCTCTCTAGGCTCTTGGTTTGACAGCTGGCTGAG 300
Qy 310 TTCAGATGCTCTGGACATCTCCAGCTGCCAGAAATCTTGTAGCTGGGCTATCCAGAAC 369
Db 301 TTCAGATGCTCTGGACATCTCCAGCTGCCAGAAATCTTGTAGCTGGGCTATCCAGAAC 360
Qy 370 CTAACCTCAAGATGAGAGTGGGAGATGCTGGCCATCATTAGGAGCTCAGGTTGTGG 429
Db 361 CTAACTTCAAGATGAGAGTGGGAGATGCTGGCCATCATTAGGAGCTCAGGTTGTGG 420
Qy 430 AGAGCTCTCTGCTGATGTGATCTGCTGGCGAGGTGACGGGGGCAATCAAGTCAGAGC 489
Db 421 AGAGCTCTCTGCTGATGTGATCTGCTGGCGAGGTGACGGGGGCAATCAAGTCAGAGC 480
Qy 490 CAGATCTGATCAATGAGGAGCCAGCTGCGCTCAGCTGCTGAGGAAATGCTGGGCCAC 549
Db 481 CAGATCTGATCAATGAGGAGCCAGCTGCGCTCAGCTGCTGAGGAAATGCTGGGCCAC 540
Qy 550 GTGGCCCAACACACACAGCTGCTCCCACTTGTGCTGAGAGACCTTGGCTTCAT 609
Db 541 GTGGCCCAACACACAGCTGCTCCCACTTGTGCTGAGAGACCTTGGCTTCAT 600
Qy 610 GCCCAGATGCGGCTGCCAGAACCTTCTCCAGGCGGCGAGCAAAAGGCTGAGAGAC 669
Db 601 GCCCAGATGCGGCTGCCAGAACCTTCTCCAGGCGGCGAGCAAAAGGCTGAGAGAC 660
Qy 670 GTGATCGGAGAGCTGCGCTTAGGAGAGTGCCTGACACCCGCTGGGCAATGATGAC 729
Db 661 GTGATCGGAGAGCTGCGCTTAGGAGAGTGCCTGACACCCGCTGGGCAATGATGAC 720
Qy 730 CGGGGCTGTGGGGGGGTGAGCCGAGAGAGTCAAGCATTTGGGGTGCAGCTCTGTGAA 789
Db 721 CGGGGCTGTGGGGGGGTGAGCCGAGAGAGTCAAGCATTTGGGGTGCAGCTCTGTGAA 780
Qy 790 CCAGGAATCTTATTTCTGAGGAACCCAGCTCTGGGCTGAGAGTTCACAGCCACAGAC 849
Db 781 CCAGGAATCTTATTTCTGAGGAACCCAGCTCTGGGCTGAGAGTTCACAGCCACAGAC 840
Qy 850 CTGCTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACGGCTGTGCTCATCTCCCTCAC 909
Db 841 CTGCTGAAGACCTTGTCCAGGCTGGCCAAAGGCAACGGCTGTGCTCATCTCCCTCAC 900
Qy 910 CAGCTGCTGTGACATCTTGAAGCTGTTGATCTGCTCTGATGAGCTGTGGAC 969
Db 901 CAGCTGCTGTGACATCTTGAAGCTGTTGATCTGCTCTGATGAGCTGTGGAC 960
Qy 970 CCATCTACTTATGAGGGGGGGCCAGACATGCTCCAGTATTCAAGCCATGCTTACCC 1029
Db 961 CCATCTACTTATGAGGGGGGGCCAGACATGCTCCAGTATTCAAGCCATGCTTACCC 1020
Qy 1030 TGTCTCTGCTACAGCAATCTGCTGATCTTATGTGACCTGACACATTTGACAGGCG 1089
Db 1021 TGTCTCTGCTACAGCAATCTGCTGATCTTATGTGACCTGACACATTTGACAGGCG 1080
Qy 1090 AGCAGAGAGCAAGAAATTTGGCCACAGGAGAAAGGCTCAGTCACTGCAAGCCCTGTTCTA 1149
Db 1081 AGCAGAGAGCAAGAAATTTGGCCACAGGAGAAAGGCTCAGTCACTGCAAGCCCTGTTCTA 1140
Qy 1150 GAAAGATGCTGATGATGATGATCTTATGAAACAGAGAGAGAGATCTGAGAG 1209
Db 1141 GAAAGATGCTGATGATGATGATCTTATGAAACAGAGAGAGAGATCTGAGAG 1200
Qy 1210 GACACCTGTGTGAAAGCAGCTGACCCACATGACCAACCACTGCCAGAGTCTACG 1269
Db 1201 GACACCTGTGTGAAAGCAGCTGACCCACATGACCAACCACTGCCAGAGTCTACG 1260
Qy 1270 AAGATGCTGGGGGGGTGAGAGATTTACAGAGCTGATCCGCTGCTCAGATTTCCAGAC 1329
Db 1261 AAGATGCTGGGGGGGTGAGAGATTTACAGAGCTGATCCGCTGCTCAGATTTCCAGAC 1320
Qy 1330 TTCCGAGACCTGCCCACCTCTCTCATCTGAGGGGAGGCTGTGATGATCAATGAGC 1389

Db 1321 TTCGAGACCTGCCACCCTCTCTCATCCATGAGGGGCGAGGCTGTGATGATCAATGAGC 1380
Qy 1390 ATCGCTCTCTATTTTGGCCATGAGAGCAATCCAGCTCTCTCATGATGATGAGCGCC 1449
Db 1381 ATCGCTCTCTATTTTGGCCATGAGAGCAATCCAGCTCTCTCATGATGATGAGCGCC 1440
Qy 1450 CTCTTGTTCATGATGCTGCTCATCCCTTTCACAGCTCATCTTGTGATGATGATGATG 1509
Db 1441 CTCTTGTTCATGATGCTGCTCATCCCTTTCACAGCTCATCTTGTGATGATGATGATG 1500
Qy 1510 TGTACTAGAGAGAGGAGCTTACTTACTATGATGAGAGAGGAGGCTGTACACACCTG 1569
Db 1501 TGTACTAGAGAGAGGAGCTTACTTACTATGATGAGAGAGGAGGCTGTACACACCTG 1560
Qy 1570 CCATATTTCTTGGCAAGATCTGAGGAGGCTTCCGAGGAGCTGCTGCTTATGATGATG 1629
Db 1561 CCATATTTCTTGGCAAGATCTGAGGAGGCTTCCGAGGAGCTGCTGCTTATGATGATG 1620
Qy 1630 TACGGGATGCCACCTACTGCTGCTGAGCAACCTGAGGCGCTGACAGCCCTTCTGCTG 1689
Db 1621 TACGGGATGCCACCTACTGCTGCTGAGCAACCTGAGGCGCTGACAGCCCTTCTGCTG 1680
Qy 1690 CACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1749
Db 1681 CACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Qy 1750 GCCCTGCTGCCACCTTCCACATGAGGCGCTTCTTCCAGCAATGCTTCAATCTCTTC 1809
Db 1741 GCCCTGCTGCCACCTTCCACATGAGGCGCTTCTTCCAGCAATGCTTCAATCTCTTC 1800
Qy 1810 TACCTGCGCGGGGCTTCAATGATTAACCTGAGAGCTGCTGAGAGTCCGCTGAT 1869
Db 1801 TACCTGCGCGGGGCTTCAATGATTAACCTGAGAGCTGCTGAGAGTCCGCTGAT 1860
Qy 1870 TCCAAAGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1929
Db 1861 TCCAAAGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
Qy 1930 AGAATTAATTAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1989
Db 1921 AGAATTAATTAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Qy 1990 AGTGCATGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2049
Db 1981 AGTGCATGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
Qy 2050 AGCGGTGCTTCAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2109
Db 2041 AGCGGTGCTTCAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
Qy 2110 CAAGACTGTGATTCAGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2169
Db 2101 CAAGACTGTGATTCAGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
Qy 2170 AACTGACCTCTCTCTGAGAGGCTTCTCTGAGGAGAGCTGAGAGCAATGACCTTACAGA 2229
Db 2161 AACTGACCTCTCTCTGAGAGGCTTCTCTGAGGAGAGCTGAGAGCAATGACCTTACAGA 2220
Qy 2230 TGTCTGACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2289
Db 2221 TGTCTGACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
Qy 2290 AGAATAAGACAGTGCAGAAAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2349
Db 2281 AGAATAAGACAGTGCAGAAAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
Qy 2350 AAAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2409
Db 2341 AAAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
Qy 2410 ATATAGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2469
Db 2401 ATATAGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2460

QY	2470	GCAGAAATGTTGGAACTCGAGGAGACAAATACAGTAGCTAGAGAGATTTGGCTTCACT	2529
Db	2461	GCAGAAATGTTGGAACTCGAGGAGACAAATACAGTAGCTAGAGAGATTTGGCTTCACT	2520
QY	2530	TCACGAGGAGCCCAACACCTCGTGAGCACCACATCATACGAAAGGACCTTAAGATGTA	2589
Db	2521	TCACGAGGAGCCCAACACCTCGTGAGCACCACATCATACGAAAGGACCTTAAGATGTA	2580
QY	2590	CCAGCAAGATGCGATCCCTCTTTTGTGTGGGTCAATGGCTCCAAAGCCAAAGCTGAA	2649
Db	2581	CCAGCAAGATGCGATCCCTCTTTTGTGTGGGTCAATGGCTCCAAAGCCAAAGCTGAA	2640
QY	2650	CAATTAATAATGATATGAGC	2669
Db	2641	CAATTAATAATGATATGAGC	2660
RESULT 3			
LOCUS	AF320294	2022 bp	mrna
DEFINITION	Homo sapiens ABCG8 (ABCG8) mRNA, complete cds.	linear	PRI 13-DEC-2000
ACCESSION	AF320294		
VERSION	AF320294.1	GI:11692801	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J., Accumulation of Dietary Cholesterol in Sltosterolemia Caused by Mutations in Adjacent ABC Transporters		
TITLE	Science (2001) In Press		
JOURNAL	2 (bases 1 to 2022)		
REFERENCE	Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J., Kvitelovich, P., Shan, B., Barnes, R. and Hobbs, H.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-NOV-2000) Molecular Genetics, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9046, USA		
JOURNAL			
FEATURES	Location/Qualifiers		
source	1..2022		
gene	/organism="Homo sapiens"		
CDS	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	1..2022		
	/gene="ABCG8"		
	1..2022		
	/note="ATP-binding cassette, subfamily G, member 8"		
	/codon_start=1		
	/product="ABCG8"		
	/protein_id="AAG40004.1"		
	/db_xref="GI:11692802"		
	/translation="MAGKAAERGLPKATPDSTGLDRFLSSSDNSLFTTSGOP		
	NTELEPLANTQVLAQVPRFEDLADFRKPMTPSPQNSCEQLVNLSEFKVSGMLA		
	TISSSGGRASLDVITGRHGKIKSGIWINQPSQVLRKCAVNRGNOLPLN		
	LVRVETLAFLAQRPRFESQADRKRVDAIARLRQCAQTRVGNMVRGLSGGR		
	RVSISGQILAMNGILIDEPSTGSDPSFAHMLVKTLRLAKGNLVLISLHOPSDI		
	ELFRLDELMTSTPTLGAOHMOVYFAIGYPCPRYSNPADFYVDLTSIDRSRRO		
	ELATREAAOSIALFLEKYRDLDDFLMKEFTKDDEPCVESTVPDPTNCLPSTKA		
	RGVADPTLIRROISNDPRDIPTLLIRAEKCLMSMTIGLITRHSIQLSPMTAA		
	LLEMLALIPNVITLDVSKYSERAMLYLEDELVTGPFERKILGELPEHCAVI		
	ITIGMPTPLMANLRGLDFLLPLMLVYPCRCIMAAALALPTFHMAFSFNAL		
	YSKFIYLAAGEMINISLMTVPAMISKVSFLMCEFGMLKIOFSRRTYKMPGNLTIAV		
	SGNFIYVLELSDYPLPYAIYIYIVIGSGFVLYVYSIRFIKQKSDQM"		
BASE COUNT	424 a	603 c	545 g
ORIGIN			450 t
Query Match	75.7%	Score 2020.4	DB 9; Length 2022;
Best Local Similarity	100.0%	Pred. No. 0;	

RESULT 4	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
AX478099	AX478099	Sequence 29 from Patent	3239 bp	GI:22217059	linear
		WO0240541.			
					PAT 12-AUG-2002

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
Source	Homo sapiens (human)							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.							
AUTHORS	1							
TITLE	Tang, Y.T., Yue, H., Nguyen, D.B., Hafalla, A.J., Elliott, V.S., Lu, Y., Wajia, N.K., Yao, M.G., Baughn, M.R., Gandhi, A.R., Ding, L., Sanjanwal, M., Ramkumar, J., Arvizu, C., Glezen, K.J., Lal, P.G., Azimzai, Y., Khan, F.A., Thangavelu, K., Thornton, M., Lu, D.A., Tiriboulet, C.M., Warren, B.A., Ison, C.H., Das, D., Raumann, B.E., Pollock, J.L. and Kearney, L.							
JOURNAL	Transporters and ion channels							
FEATURES	Patent: WO 0240541-A 29 23-May-2002; Incyte Genomics, Inc. (US)							
Source	Location/Qualifiers							
	1..3239							
	/organism="Homo sapiens"							
	/mol_type="genomic DNA"							
	/db_xref="taxon:9606"							
	/note="Incyte ID No: 6585710CB1"							
BASE COUNT	784 a 822 c 796 g 837 t							
ORIGIN								
Query Match	63.0%; Score 1680.6; DB 6; Length 3239;							
Best Local Similarity	99.8%; Pred. No. 0;							
Matches 1683; Conservative	0; Mismatches 4; Indels 0; Gaps 0;							
983	GGGGGGCCGAGACATGGTCCAGTCTTTACAGCCATCGGCTACCCCTCTCTCCCTACA	1044						
12	GGGGGGCCGAGACATGGTCCAGTCTTTACAGCCATCGGCTACCCCTCTCTCCCTACA	71						
1043	GCAATCCGTCGACTTCTATGTGACCTGACCGACGATTTACAGCGGACAGAGAGCAGG	1107						
72	GCAATCCGTCGACTTCTATGTGACCTGACCGACGATTTACAGCGGACAGAGAGCAGG	131						
1103	AATTGGCCACAGGAGAGAGCTCAGTCTGCGACCCCTGTTCTAGAAAAAGTGCCTG	1167						
132	AATTGGCCACAGGAGAGAGCTCAGTCTGCGACCCCTGTTCTAGAAAAAGTGCCTG	191						
1163	ACTTAGATGATCTTCTATGGAAGACAGAGAGAGATCTTAGAGAGACACCTGTGCG	1222						
192	ACTTAGATGATCTTCTATGGAAGACAGAGAGATCTTAGAGAGACACCTGTGTGG	251						
1223	AAAGCAGCGTACCCCTACATGACACCAACTCCCTCCAGTCTCTACGAAGATCCTGGGG	1283						
252	AAAGCAGCGTACCCCTACATGACACCAACTCCCTCCAGTCTCTACGAAGATCCTGGGG	311						
1283	CGGTGACAGCATTTACGACGCTGATCCGCTCAGATTTCCAAACGACTTCCGAGACCTGC	1342						
312	CGGTGACAGCATTTACGACGCTGATCCGCTCAGATTTCCAAACGACTTCCGAGACCTGC	371						
1343	CCACCCTTCATCATGAGGGGGGGGGGGCTGTCGATGATGCAATGACATCGGCTTCCT	1407						
372	CCACCCTTCATCATGAGGGGGGGGGGGCTGTCGATGATGCAATGACATCGGCTTCCT	431						
1403	ATTTTGGCCATGGAGAGCATCCAGCTCTCTTCATGATGATACAGCGCCCTCTGTTCATGA	1467						
432	ATTTTGGCCATGGAGAGCATCCAGCTCTCTTCATGATGATACAGCGCCCTCTGTTCATGA	491						
1463	TCGGTGCCTCATCCCTTTCAAGCTCATTTGATGATGATGATGATGATGATGATGATGATG	1522						
492	TCGGTGCCTCATCCCTTTCAAGCTCATTTGATGATGATGATGATGATGATGATGATGATG	551						
1523	GGGGAATGCTTACTATGAACTGGAAGCGGGGGTACACAGAGGTCATATTTCTTGG	1588						
552	GGGGAATGCTTACTATGAACTGGAAGCGGGGGTACACAGAGGTCATATTTCTTGG	611						
1583	CCAAGATCTTCGGGGAGCTTCGGAGCAGTGTGCTCATCATCATCATCATCATCATCATCAT	1642						
612	CCAAGATCTTCGGGGAGCTTCGGAGCAGTGTGCTCATCATCATCATCATCATCATCATCAT	671						
1643	CTTACTGGCTGGCAACCTGAGCCGAGGCTCAGCCCTTCTGCTGCACTTCTGCTGTGG	1702						

Db 672 CCTACTGGCTGGCCAACTGAGCGCCCTCAAGCCCTCTCTGCTGACCTTCCGCTGG 731
QY 1703 TGTGGCTGGTGGTCTTCTGTTGACAGATTATGAGCCCTGCGCCGCCGCCCTGCTCCCA 1762
Db 732 TGTGGCTGGTGGTCTTCTGTTGACAGATTATGAGCCCTGCGCCGCCGCCCTGCTCCCA 791
QY 1763 CCTTCACAGTGGCT 1822
Db 792 CCTTCACAGTGGCT 851
QY 1823 GCTTCATGATATAAATTGAGCGCCTGTGAGAGTGCCTGCTCTCTCTCTCTCTCTCTCT 1882
Db 852 GCTTCATGATATAAATTGAGCGCCTGTGAGAGTGCCTGCTCTCTCTCTCTCTCTCTCT 911
QY 1883 TCTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1942
Db 912 TCTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 971
QY 1943 TGCCTCTCGGGAACTCACCATCCGCGTCTCAGAGATAAATAATCTCAGTGCATGGAGC 2002
Db 972 TGCCTCTCGGGAACTCACCATCCGCGTCTCAGAGATAAATAATCTCAGTGCATGGAGC 1031
QY 2003 TGGACTGCTACCT 2062
Db 1032 TGGACTGCTACCT 1091
QY 2063 TGGACTGCTACCT 2122
Db 1092 TGGACTGCTACCT 1151
QY 2123 TCAAGCGCAGAGCTGTGCGCCCTGTGTGGGGAGCTGAGACAGACCTTCACTGACCTCCT 2182
Db 1152 TCAAGCGCAGAGCTGTGCGCCCTGTGTGGGGAGCTGAGACAGACCTTCACTGACCTCCT 1211
QY 2183 CCTCAGAGAGCCCT 2242
Db 1212 CCTCAGAGAGCCCT 1271
QY 2243 CCGGCGCAGAGCTGTGAGTGGGACAGACAGCAGCAGAGTGGAGTGAAGTAAAGCAG 2302
Db 1272 CCGGCGCAGAGCTGTGAGTGGGACAGACAGCAGCAGAGTGGAGTGAAGTAAAGCAG 1331
QY 2303 TCGAAGAGGATTTCTGCTCACTGTGAGAGAGTGCATGCTGGAGAGAAACCTGACCTC 2362
Db 1332 TCGAAGAGGATTTCTGCTCACTGTGAGAGAGTGCATGCTGGAGAGAAACCTGACCTC 1391
QY 2363 GGTGGGACCTACACAGTGTCTAATTTATTTCTCTTTGATGATGATTTATATAGCAACTC 2422
Db 1392 GGTGGGACCTACACAGTGTCTAATTTATTTCTCTTTGATGATGATTTATATAGCAACTC 1451
QY 2423 GATATAGGATGGAGGAGCAATAGATGATGATGATGATGATGATGATGATGATGATG 2482
Db 1452 GATATAGGATGGAGGAGCAATAGATGATGATGATGATGATGATGATGATGATGATG 1511
QY 2483 GAACTGAGAGGAGCAATAGATGATGATGATGATGATGATGATGATGATGATGATG 2542
Db 1512 GAACTGAGAGGAGCAATAGATGATGATGATGATGATGATGATGATGATGATGATG 1571
QY 2543 CACTCGGTGGTGGAGCAGATCAATACAGAAAGTGAAGTGAAGTGAAGTGAAGTGA 2602
Db 1572 CACTCGGTGGTGGAGCAGATCAATACAGAAAGTGAAGTGAAGTGAAGTGAAGTGA 1631
QY 2603 ATCCCT 2662
Db 1632 ATCCCT 1691
QY 2663 ATTCAGC 2669
Db 1692 ATTCAGC 1698

RESULT 5
AF351785

LOCUS AF351785 4829 bp mRNA linear ROD 26-AUG-2002
DEFINITION Rattus norvegicus sterol-2 (Abcg8) mRNA, complete cds.
ACCESSION AF351785
VERSION AF351785.2 GI:22477145
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4829)
AUTHORS Lu, K., Lee, M. H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
Ose, L., Stalenhoef, A. F., Miettinen, T., Bjorkhem, I., Bruckert, E.,
Pandya, A., Brewer, H. B. Jr., Salen, G., Dean, M., Stivastava, A. and
Patel, S. B.
Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterol-1 and
sterol-2, encoded by ABCG5 and ABCG8, respectively
Am. J. Hum. Genet. 69 (2), 278-290 (2001)
MEDLINE 21344600
PUBMED 11452359
REFERENCE 2 (bases 1 to 4829)
AUTHORS Lu, K., Yu, H., Lee, M. and Patel, S. B.
Molecular cloning, genomic structure, and characterization of novel
mouse head-to-head tandem ABC transporters
Unpublished
REFERENCE 3 (bases 1 to 4829)
AUTHORS Lu, K., Lee, M. and Patel, S. B.
Direct Submission
TITLE Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
Sc, STB 541, Charleston, SC 29403, USA
4 (bases 1 to 4829)
AUTHORS Lu, K., Yu, H., Lee, M. and Patel, S. B.
Direct Submission
TITLE Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
Sc, STB 541, Charleston, SC 29403, USA
COMMENT Sequence update by submitter
On Aug 26, 2002 this sequence version replaced gi:15148516.
FEATURES
location/Qualifiers
1..4829
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1..4829
/gene="Abcg8"
11..2129
/gene="Abcg8"
/codon_start=1
/product="sterol-2"
/protein_id="AAK84831.2"
/db_xref="GI:22477146"
/translation="MAEKTKETPLNGVTVLQDASSLQDSVFSSSDNSLYTYSQGS
NTLEVEDLYQVDMASQVPEFQLOAFKLPWRBRSQDSQDSDLEIRINSEFVRSGOMIA
IIGSACGRATLIDVTGRDHGSKMSGQVINGQSTPOLIOKCAVHVAQOOLPN
LVRETLFTLAORHLRPFESQARQDRKREVDVALELRQCAVNTVRYGVSGGER
RRYSIGVOLLNMGITLILDEPSGILSFVFAHNVRSLRAKNNRVLISLRPSDI
FRFDVLMTSTPTIYLVQAQMVOTFTSISGICPRYSHPADFYVLDNISDRSEQ
EVATMEKARLUALFEKQVQEDFLMKAEKASLDGTAVASQTLVDQDNCTGAELP
GMIOQFTLIRQISNDFDPLFIHGAELACMSLIGFLYGGADPLSEFMDMAELP
LPMIGALIPENVILDVSKSRSLSLYELEDGLTAGPYFAKVIQGLPEHCAVYI
IKGMPYLVNLIRPGPELFLHFMILMLVYPCCRMLAASAMLPETHSSRCNALLY
NSPYLTFAGMININMNIYPAVSTKSEFLKPECSGLMOIQPNHITTTTIGNLTSVP
GDAAVTAHMLNSHPLAIVLIVIGISCGFLSYLSLRFIKOKSIDW"

BASE COUNT 1105 a 1292 c 1214 g 1218 t

Query Match 56.9%; Score 1518.6; DB 10; Length 4829;
Best local similarity 77.6%; Pred. No. 0;
Matches 1973; Conservative 0; Mismatches 514; Indels 56; Gaps 9;

99 CATGGCGGGAAGGCGGAGAGAGAGAGGCTGCCGAAGGCGCCACTCCCGAGGATAC 158
Db 110 CATGGCTAGAGAACCAAGAGAGAGACCCAGCTGTGGAACGGAGCTACTACAGAGATGC 169
Oy 159 CTCGGGCTCCAGAGATAGATTGTCTCTCTGAAAGTGAACAAGACCTGTACTTACCTA 218
Db 170 TTCAAGCCTCCAGAGACAGCGTGTCTCTCTGAAAGTGAACAAGACCTGTACTTACCTA 229
Oy 219 CAGTGGCCAGCCCAACACCTGTGAGAGAGAGAGACCTCAACTACAGTGTGAGCTGCCTC 278
Db 230 CAGTGGCTAGTCAACACTGTGAGAGTCAAGATCTACACTACAGAGTGTGAGAGCTGCCTC 289
Oy 279 TCAGTCCCTGTGTGAGAGAGTGTGAGAGAGCTGTGAGAGAGCTGTGAGAGTGTGAGAG 338
Db 290 TCAGTCCCTGTGTGAGAGAGTGTGAGAGAGCTGTGAGAGAGCTGTGAGAGTGTGAGAG 349
Oy 339 CCAGAAATCTGTGAG 398
Db 350 CCAGAGACTGTGTGAT 409
Oy 399 GCTGGCCATCATPAGGAG 458
Db 410 GCTGGCTATCATPAGGAG 469
Oy 459 CCGAG 518
Db 470 CAG 529
Oy 519 GCTCATAGTGTGAG 578
Db 530 GCTCATAGTGTGAG 589
Oy 579 CTTACATGTGTGAG 638
Db 590 TCTCATGTGTGAG 649
Oy 639 CCAG 698
Db 650 TCAG 709
Oy 699 CGCTGAG 758
Db 710 CGCGAG 769
Oy 759 AGTACAGATGTGTGAG 818
Db 770 AGTACAGATGTGTGAG 829
Oy 819 CTTGAG 878
Db 830 TTTGAG 889
Oy 879 AGGAG 938
Db 890 AGGAG 949
Oy 939 TGAATGTGTGTCTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
Db 950 TGACCTGTGTCTCTGATGAG 1009
Oy 999 GGTGAGATTTTACAG 1058
Db 1010 GGTGAGATTTTACAG 1069
Oy 1059 CTATGTGAG 1118
Db 1070 CTATGTGAG 1129
Oy 1119 GAAAGCTGAGTATGAG 1178
Db 1130 GAAAGCTGAGTATGAG 1189
Oy 1179 ATGGAAG 1238

Db 1190 GTGGAAG 1249
Oy 1239 ACTAG 1298
Db 1250 ACAG 1306
Oy 1299 GAGCCTGATCCGTGTGAG 1358
Db 1307 CACCTGATCCGTGTGAG 1366
Oy 1359 TGGGCGAG 1418
Db 1367 TGGAG 1426
Oy 1419 CATGAG 1478
Db 1427 CAAAG 1486
Oy 1479 TTTCAAG 1538
Db 1487 TTTCAAG 1546
Oy 1539 TGAAG 1598
Db 1547 TGAAG 1606
Oy 1599 GCTTCCGAG 1658
Db 1607 GCTTCCGAG 1666
Oy 1659 CTTGAG 1718
Db 1667 CTTGAG 1726
Oy 1719 CTTGAG 1778
Db 1727 CTTGAG 1786
Oy 1779 CTTGAG 1838
Db 1787 CTTGAG 1846
Oy 1839 GAG 1898
Db 1847 GAG 1906
Oy 1899 TGAAG 1958
Db 1907 CTTGAG 1966
Oy 1959 CACCTGAG 2018
Db 1967 CACCTGAG 2026
Oy 2019 CTTGAG 2078
Db 2027 TTTGAG 2086
Oy 2079 GTCCTTGAAGTGTATCAAG 2138
Db 2087 GTCCTTGAAGTGTATCAAG 2143
Oy 2139 CCGGCTGTGGGAG 2198
Db 2144 TCCAGTGTGGGAG 2192
Oy 2199 CTTGAG 2258
Db 2193 GAGTGTGGGAG 2245
Oy 2259 CAGTGTGGGAG 2318

Db	2246	CAGTGGCAGACGTCGCCACAGAGATGGAGTAGAATTAAAGCACTGTAGAG--TTTCTG	2303
Oy	2319	CTCACTGGCAGAGACTCCGATGACTGGAGAAAACTGCACCTCGTGTCACCTACAACG	2378
Db	2304	CTCACAGGCGCTGGGGCTGTGTG-----AAACAGTAGACTTCGTGAACCTGTAAACG	2350
Oy	2379	TTCGTAATTTATTTCCCTTTTGATATGCAATTTATATAGCACTCATATATAGATGGGAC	2438
Db	2351	TTGCTCANTCATTT-----TATCTCATATAAAAACCACAGTAGATGGAAATGGGAAC	2402
Oy	2439	AAACTAGAATGAATTTGGGTAGCTAGACTGACAGCAATTTGTGGAACTGTGAGGAGAAC	2498
Db	2403	CAATATATTATGAANTGTAGTAGTAGCTATGACAGAGACTGTGTGAACCCGAGAGCATR	2462
Oy	2489	ATACACGTAGCTACGACATTTGGCTTCATCTCCAGGGGCCCACTCCCTGTGTGACC	2558
Db	2463	GTCGTCTGTACCAAAACATTTAGGTCTTCTCCACCA-----ATCTACCCCTGTATTAACG	2515
Oy	2559	ACCATCAATPACAGAAAGTACCTTAAGATAGTACCGACAGATGATCCATCCTCTTTTGTG	2618
Db	2516	GCTCCCGATACAGAGGGTGACCTAAACGTACTAG-AAAATGCTCTCTTATCTCTGTG	2573
Oy	2619	TGGGCTCATTTGGCTCCAAAACCC	2641
Db	2574	TGGCTCCATGTGACCTCCAGATCTC	2596
RESULT 6			
AF324495			
LOCUS	AF324495	3674 bp	mRNA linear ROD 07-AUG-2001
DEFINITION	Mus musculus sterolin-2 (Abcg8) mRNA, complete cds.		
ACCESSION	AF324495		
VERSION	AF324495.1	GI:15088541	
KEYWORDS			
SOURCE			
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3674)		
AUTHORS	Lu,K., Lee,M.H., Hazard,S., Brooks-Wilson,A., Hidak,H., Kojima,H., Ose,I., Stalenhoef,A.F., Meitinen,T., Bjorkhem,I., Bruckert,E., Pandya,A., Brewer,H.B. Jr., Salen,G., Dean,M., Sivasubava,A. and Patel,S.B.		
TITLE	Two genes that map to the STRL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by Abcg5 and Abcg8, respectively		
JOURNAL MEDLINE	Am. J. Hum. Genet. 69 (2), 278-290 (2001)		
PUBMED	21344600		
REFERENCE	11452359		
AUTHORS	2 (bases 1 to 3674)		
TITLE	Lu,K., Lee,M.-H. and Patel,S.B.		
JOURNAL	Direct Submission		
FEATURES	Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty Street, STB541, Charleston, SC 29403, USA		
SOURCE	Location/Qualifiers		
	1..3674	/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6"	
		/db_xref="taxon:10090"	
		/tissue_type="liver"	
		1..3674	
		/gene="Abcg8"	
		102..2123	
		/gene="Abcg8"	
		/note="Abcg8"	
		/codon_start=1	
		/product="sterolin-2"	
		/protein_id="AK84079.1"	
		/db_xref="GI:15088542"	
		/translation="MAEKKEETOLINGTVADASOGIDSLFSSSDNSLYTYSGQ SNLEVRDLITQVDIASQYPMFPQLAQFKLPWSHSSQDSCELGIRNLSTKVSQML	
CDS			
gene			

Query Match	Similarity	Score	Length
Best Local	77.0%	1511.6	3674
Matches 1965;	Conservative 0;	Mismatches 534;	Indels 53; Gaps 8;
QY	99	CATGGCCGGGAAGGCGGACAGAGAGAGAGAGGAGGCTGCCAAGGGGCCACTCCCAAGATAC	158
Db	101	CATGGCTGTAGAAACCAAGAGAGAGACCACCTGTGAAATGGGACTGACTACTCAGAGATGC	160
QY	159	CTC---GGGCTCCAGGATAGATTGTTCCCTCGAAGATGACAAACACCTGTAATCCAC	215
Db	161	TTCCGAGAGGCTCCAGAGACACTTGTCTCTCCGAAAGTACAAACACTGTGTACTTAC	220
QY	216	CTACAGTGGCCAGCCCAACACCTGGAGAGTCAGAGACCTCAACTACAGTGTGACCTTGC	275
Db	221	CTACAGTGGTCAATCCAAACACTGTGAGAGTCAGAGATCTCACTACAGGTGGACATGCG	280
QY	276	CTCCAGTCCCTGTGGTTTGAGCAGTGGCTCAATTCAAGTGGCTGTGACATCTCCAG	335
Db	281	CTCTAGGTGCTTGTGTTTGAGCAGTGGCTCACTTCAAGATACCTGTGAGGTCTCATAG	340
QY	336	CTCCAGAAATCTTGTAGCTGGGCAATCCAGAACCTTAAAGCTTCAAAGTGAAGAGGGCA	395
Db	341	CAGCCAAAGTCTCTGTAGGTGGGCAATCCAGAACTTAAAGCTTCAAAGTGAAGATGGACA	400
QY	396	GATCTGGCCATCATAGGAGCTCAGGTTGTGGAGAGCCTCTTGTAGATGTGATCAC	455
Db	401	GATCTGGCCATCATAGGAGCTCAGGCTGGGGAGAGCCCACTACTCGACGTGATCAC	460
QY	456	TGGCCGAGGTACAGGGCGGCAAGATCAAGTACAGGCAATCTGATTCATAGGGACGCCAG	515
Db	461	AGGCAAGAGGCCACGCTGACAGATGAATACAGCAAAATTTGGATTAATGGGCAACCCAG	520
QY	516	CTGCCCTCAGTGTGAGAAAGTGTGTGGCCCAAGTGCAGCAGACAACAAGCTGCTGCC	575
Db	521	TAGCCCTCAGTGTGAGAAAGTGTGTGGCCCAAGTGCAGCAGATACACAACCTGCTGCC	580
QY	576	CAACTTACATGTGGGAAGACCTTGGCCCTTCATTGCCAGATGCGGCTGCCCAACCTT	635
Db	581	CAACTTACATGTGGGAAGACCTTGGCCCTTCATTGCCCAATGCGGCTGCCCAAGACCTT	640
QY	636	CTCCAGAGGCCAGGCTGACAAAAGGGGAGAGAGGTATGGCGGAGTGGGGCTTAGGA	695
Db	641	CTCCAGAGGCCAGGCTGACAAAAGGGGAGAGAGGTATGGCGGAGTGGGGCTTAGGA	700
QY	696	GTCGCTGACACCCGCTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAAGCCAG	755
Db	701	GTCGCTGACACCCAGGAGTGGGCAACACGTATGTACGTGGGGTTGTGCGGGGGTGAAGCCG	760
QY	756	GAGAGTGAATGGGGTGTGAGTCCCTGTGGGAACCCAGGAATCTTATTTCTCGACAGAC	815
Db	761	ACGGTGAATGGGGTGTGAGTCCCTGTGGGAACCCAGGAATCTTATTTCTGAATGA	820
QY	816	CACCTCTGGGCTGACAGCTTACAGGCCCAACACTGTGTAAAGACTTGTCCAGAGCTGGC	875
Db	821	CACCTCTGGGCTGACAGCTTACAGGCCCAACACTGTGTGTACAAACCTTGTCCCGGCTGGC	880
QY	876	CAAAAGCAACGGCGTGGTGTCAATCTCCCTCCACACACCTCGCTGTGACATCTTCAAGCT	935
Db	881	CAAGGGCAACAGGCTGGTGTCTCAATCTCCCTCCACACAGCTCGCTGTGACATCTTCAAGCT	940

OY 936 GTTATGCTGTCCTGATGACGCTGGACACCCCATCTACTAGGGGGGCCACGA 995
 DB 941 ATTTACCTGCTGTTGATGATCTGGACACCCCTATCTACCTGGGGGGGCCACGA 1000
 OY 996 CATGCTCAGATTTCACAGCCTGGGCTACCCCTGCTGCTGCTACCAATCCGCTGA 1055
 DB 1001 ATGCTGACGATCTTACATCTGATGGCCCTTGTCTGCTATACCAACCTGGCGGA 1060
 OY 1056 CTCTATGTGACCTGACACAGATTGACAGGGGACAGAGAGAGAGAAATGGCCACAG 1115
 DB 1061 CTCTACGCTGACCTGACACAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
 OY 1116 GGAGAGGCTCAGTACTGACAGCCCTTCTTCTGAAAAAGTGGCTACTAGATGATT 1175
 DB 1121 GGAGAGGCTCAGTACTGACAGCCCTTCTTCTGAAAAAGTGGCTACTAGATGATT 1180
 OY 1176 TCTATGAG 1235
 DB 1181 TCTGAG 1240
 OY 1236 CCCACTAGACACCACTGCTCCGAGTCTACAGAGATGCTGGGGGGCTGACAGATT 1295
 DB 1241 CACACAGAGACACTGCTG---TGGGACTGCTGTTGAGCTGCGCGGAGATGAGAGATT 1297
 OY 1296 TACGAGCTGATCCGCTGCTCAGATTTCACAGACTCCGAGACTGCCACCTCTCAT 1355
 DB 1298 TTCCACCTGATCCGCTGCTCAGATTTCACAGACTCCGAGACTGCCACCTCTCAT 1357
 OY 1356 CCATGGGGGGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1415
 DB 1358 TCTATGAG 1417
 OY 1416 GAGACCTCAGCTCTCTCTCATGATGATGATGATGATGATGATGATGATGATGATG 1475
 DB 1418 GGCACAGAGAGCTCTCTCTCATGATGATGATGATGATGATGATGATGATGATGATG 1477
 OY 1476 CCCCTTCACTCATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1535
 DB 1478 TCTCTTCAATGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1537
 OY 1536 CTATGACTGAG 1595
 DB 1538 CTATGACTGAG 1597
 OY 1596 GGAGCTTCCGAGAGACTGCTGCTCATCATCATCATCATCATCATCATCATCATCATG 1655
 DB 1598 AGAATGCTCGGAG 1657
 OY 1656 CAACCTAG 1715
 DB 1658 AACCCTGCGGCGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1717
 OY 1716 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1775
 DB 1718 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1777
 OY 1776 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1835
 DB 1778 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1837
 OY 1836 CTGAG 1895
 DB 1838 CTGAG 1897
 OY 1896 TTTTGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1955
 DB 1898 CTCTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1957
 OY 1956 CTCTACCACTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2015
 DB 1958 CTCTACCACTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2017
 OY 2016 TCTCTAGGCACTACTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTA 2075

DB 2018 ACTATATCGATCTACTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTA 2077
 OY 2076 CGTGTCTTAAAGTCTCATCAAAACAAACAAAGTCAAAAGTCAAAAGTCAAAAGTCAAA 2135
 DB 2078 TCTATCTTGAAGCTCATCAAAACAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 2137
 OY 2136 CTGCGGCTGGTGGGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2195
 DB 2138 CTCACTGGCGG-----GACCTTTTCCCGGGGCTGGCCACCCAGAGAGAGAGAG 2185
 OY 2196 TTCTGGGGAG 2255
 DB 2186 GGAGCTGGGGAG 2238
 OY 2256 CTGAGTGGGAG 2315
 DB 2239 CTGAGTGGGAG 2298
 OY 2316 CTGCTACTGAG 2375
 DB 2299 CTGCTCCAG 2343
 OY 2376 AGTGTCTTAAATTTCTCTTTTGTATGATGATGATGATGATGATGATGATGATGATG 2435
 DB 2344 AGTGTCTTAAATTTCTCTTTTGTATGATGATGATGATGATGATGATGATGATGATG 2395
 OY 2436 AGCAACAG 2495
 DB 2396 AACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2455
 OY 2496 ACAATTAACAG 2555
 DB 2456 ATAGTGTCTTAAATTTCTCTTTTGTATGATGATGATGATGATGATGATGATGATGATG 2511
 OY 2556 GCCAACCATTAACAG 2614
 DB 2512 GCCAACCATTAACAG 2571
 OY 2615 TGTGTGGGTCATGGGCTCCCAAAAGCCAGCT 2646
 DB 2572 TGTGTGGGTCATGGGCTCCCAAAAGCCAGCT 2603

RESULT 7
 AY196216
 LOCUS
 DEFINITION
 Mus musculus strain PERA/EI ATP-binding cassette sub-family G
 member 8 (Abcg8) mRNA, complete cds.
 ACCESSION
 AY196216
 VERSION
 AY196216.1 GI:31322261
 KEYWORDS
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2284)
 Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and
 Paigen, B.
 Primary Roles of Fxr and Abcg5/Abcg8 in cholesterol gallstone
 Susceptibility: Evidence from a Cross of PERA/EI and I/Jn Inbred
 Mice
 Unpublished
 2 (bases 1 to 2284)
 Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
 Direct Submission
 Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street,
 Bar Harbor, ME 04609, USA
 FEATURES
 source
 1. 2284
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="PERA/EI"

QY	1779	CTTCTGAGCAATGCCCTTCAACACTCTTCTTACTCTGCGGGGGCTTCAATGATAACTT	1838
Db	1778	CTTCTTGTGCATATGCCCTTCAACACTCTTCTTACTCTGCGGGGGCTTCAATGATAACTT	1837
QY	1839	GAGCAGCCTGTGGACAGAGCCCGGGTGAATTTCCAAAGTGTCTTCCCGGGGTGTGT	1838
Db	1838	GGACACACTGTGTGATAGTGTCTGTATGATTTCCAAAGTGTCTTCCCGGGGTGTGT	1837
QY	1899	TGAAGGCTGTGAAGATTCAGTTTCAGCAGAGAACTATATATAATGTGCTTCGGGAACCT	1958
Db	1898	CTCGGGGCTGTATGAGATTCATTAATTTAATGACACACTTTACACACACATAATGGCAACTT	1957
QY	1959	CACCATGCGGCTTCAGAGATTAATATCTCAGTGTCCATGTGAGTGTGACTGTACCTCT	2018
Db	1958	CACCTTTCCTCCATCTCGGAGACAGATGATGATGACCTGCACTGCACTGCACTTCACCT	2017
QY	2019	CTACGCGCATTCACCATCGTATGTCATGGCTCAGCGGGCTTCATATGGCTCTGTACTACT	2078
Db	2018	CTATGCGATTCATCTCATGTATGTATCGGCTATGATGAGGCTTCTGTCTGTACTACT	2077
QY	2079	GTCCTTAAGTTTCATCAACAGAAACCAAGTCAAGTGTGATTCACAGCCAGACGTCTG	2138
Db	2078	ATCTTGAAGTGTATCAAAAGAGTGTATTCATCAAGATGTGTGATTCACAGCTTGTCTC	2137
QY	2139	CCGCTGTGTGGGGAGCTTGACAGACCCCTTCACTGCATCCCTCTCAGAGGCCCTTTC	2198
Db	2138	ACTGTGCGG-----GACCTTTTTCCTCGGGCTGGCCACCCAGAGGAGCCGGGA	2185
QY	2199	CTGGGGACAGTGAAGCAATATGACCTTACAGATGTCTCAGTATACATCCGGCCAGGGTCTG	2258
Db	2186	CTGGGGACAGAGGCTCACACAGATTCATCAG-----GCAGCACCCACCTTGTGTGCTG	2238
QY	2259	CATGGCAGACAGACCAGCCAGATGATGGCAGTATGATTAAGCACT	2303
Db	2239	CAGTGGCAGAGCTACGCCACAGATGGCAGTATGATTAAGCACT	2283
RESULT 8			
LOCUS	AY196215		
DEFINITION	AY196215	2285 bp mRNA linear ROD 01-JUN-2003	
ACCESSION	AY196215	Mus musculus strain I/LmJ ATP-binding cassette sub-family G member	
VERSION	AY196215	8 (Abc98) mRNA, complete cds.	
KEYWORDS	AY196215.1	GI:31322259	
SOURCE			
ORGANISM		Mus musculus (house mouse)	
REFERENCE			
AUTHORS		Eubayrola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and Paigen, B.	
JOURNAL		Primary Roles of FKR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/EI and I/LmJ Inbred Mice	
AUTHORS		Unpublished	
TITLE		2 (bases 1 to 2285)	
JOURNAL		Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.	
FEATURES		Direct Submission	
SOURCE		Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA	
		Location/Qualifiers	
		1..2285	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="I/LmJ"	
		/db_xref="taxon:10090"	
		/chromosome="17"	
		/map="55 CM"	
		/sex="male"	
		/tissue_type="liver"	
		1..2285	
		/gene="Abcg8"	
gene			

[illegible]

Db 761 AGTACGATTTGGGTGACGCTCCGTGGAACCCAGAACTCTATCTGTGATGAACCCAC 820
QY 819 CTCTGGGCTGGACAGCTTCACAGCCCAACCTGTGTAGAACCTTGTCCAGGCTGGCAA 878
Db 821 TTCTGGCTTGGACAGCTTCACAGCCCAACCTGTGTAGAACCTTGTCCAGGCTGGCAA 880
QY 879 AGGCAACCGGCTGGTGTCTCTCCCTCCAGCCGCTGTGACATCTTCAGGCTGT 938
Db 881 GGGCAACAGGCTGTGTCTCTCCCTCCAGCCGCTGTGACATCTTCAGGCTGT 940
QY 939 TGATCTGTCTCTGTATGACAGCTGTGACACCCCTCTACTTACTTGGGCGGCGGACAT 998
Db 941 TGACCTGTCTCTGTATGACATCTGTGACACCCCTCTACTTACTTGGGCGGCGGACAT 1000
QY 999 GGTTCAGATTTATACAGACCTGGGCTACCCCTCTCTGCTACAGACATCTGTACTT 1058
Db 1001 GGTTCAGATTTATACATCAATTTGGCCACCTTCTCTGCTATAGCAACCTTCAGACTT 1060
QY 1059 CTATGTGACCTGACAGCATTTGACAGGCGCAGACAGGACAGAAATTTGGCCACAGGA 1118
Db 1061 CTACGTGACATTTACAGCATTTGACAGGCGCAGACAGGAGGAGTGCCACCTGGGA 1120
QY 1119 GAAAGCTCAGTCTGACAGCCCTGTTTCTAGAAAAAGTCCGTGACTTATGACTTTCT 1178
Db 1121 GAAAGCTCAGTCTTGTGACAGCCCTGTTTCTAGAAAAAGTCAAGGCTTGTATGACTTTCT 1180
QY 1179 ATGGAAGCAGACAGCAAGATTTTACAGAGACACCTGTGTGAAAAAGCAGCTGACCC 1238
Db 1181 GTGGAAGCTGAGCAAGCAAGATTTTCAACACACACACACAGCTGTGACCTGTAC 1240
QY 1239 ACTAGACACCAACTGCTCCGAGTCTCTGAAAGATGCTGGGGGCGGTGACAGCTTAC 1298
Db 1241 ACAGGACACTGACTG---TGGGACTCTCTGTGAGCTGCCGGGATATAGAGAGTTTC 1297
QY 1299 GACGCTGATCCGTGTCAGATTTTCAACAGACTTCCAGACCTGCCACCTCTCATCA 1358
Db 1298 CACCTGATCCGTGTCAGATTTTCAACAGACTTCCGGGAGCTGCCACCTCTCATTTCA 1357
QY 1359 TGGGGGGAGGCTGTCTATGATGATGACATGACATGCGCTCTCTATTTTGGCCAGGAG 1418
Db 1358 TGGGGGGAGGCTGTCTATGATGATGACATGACATGCGCTCTCTATTTTGGCCAGGAG 1417
QY 1419 CATCCAGCTCTCTCTATGATGATGATGACAGCCGCTCTGTATGATGAGGCTCTATCC 1478
Db 1418 CAGGAGCTCTCTCTATGATGATGATGACAGCCGCTCTCTATGATGAGGCTCTATCC 1477
QY 1479 TTTCAACGTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1538
Db 1478 TTTCAATGTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1537
QY 1539 TGAACGTGGAAGAGGGCTGTACACACAGCTGTGTCATATTTCTTTGCCAAGATCTCGGGA 1598
Db 1538 TGAAGCTGGAAGAGGGCTGTACACACAGCTGTGTCATATTTCTTTGCCAAGATCTCGGGA 1597
QY 1599 GCTTCGAGAGCACTGTGCTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1658
Db 1598 ATTGCGGAGCACTGTGCTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1657
QY 1659 CCTGAGGAGGAGGCTGTACAGCCCTCTCTGCTGACATTTCTGCTGTGCTGTGCTGT 1718
Db 1658 CCTGAGGAGGAGGCTGTACAGCCCTCTCTGCTGACATTTCTGCTGTGCTGTGCTGT 1717
QY 1719 CTGTTGCAAGATTTATGCGCTGTGCGCGCGCGCGCTGCTGCCACCTTCACATGAGCTC 1778
Db 1718 CTGTTGCAAGATTTATGCGCTGTGCGCGCGCGCGCTGCTGCCACCTTCACATGAGCTC 1777
QY 1779 CTCTTTCAGCAATGCGCTTACACTCTCTCTACTCTGCGCGGCTTCATATTAACCT 1838
Db 1778 CTCTTTCAGCAATGCGCTTACACTCTCTCTACTCTGCGCGGCTTCATATTAACCT 1837
QY 1839 GAGCAGCCTGTGAGAGTGGCGCGCTGTGATTTTCAAAAGTCTCTCTGCGGCTGT 1898
Db 1838 GGACAAAGCTGTGAGAGTGGCGCGCTGTGATTTTCAAAAGTCTCTCTGCGGCTGT 1897

QY 1899 TGAAGGCTGTATAGATTTAGTTACAGCAGAACTATTAATGCTCTCGGAACT 1958
Db 1898 CTCGGGCTGTATGAGATTTATTTATGAGACACTTTACACACAAATTCGGAACCT 1957
QY 1959 CACCATTGCGGCTGTACAGATTAATCTCTAGTGCATGTGAGCTGTGATCTCT 2018
Db 1958 CACCTTCTCCATCTCTGAGACAGATGATGATGATGATGATGATGATGATGATGAT 2017
QY 2019 CTACGCATCTACTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2078
Db 2078 CTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2077
QY 2079 GTCTTAAAGTTATCAACAGAAACCAAGTCAAGATGATGATGATGATGATGATGAT 2138
Db 2078 ATCTTGAAGCTCATCAACAGAAACCAAGTCAAGATGATGATGATGATGATGATGAT 2137
QY 2139 CCGGCTGTGTGGGAGCTGTGAGCAGACCTTCAACTGACTCTCTCTGAGAGCCCTTC 2198
Db 2138 ACTGGCG-----GGACCTTCTCCGCGGCTGTGCGCCACCCAGAGAGGCGGA 2184
QY 2199 CTGGGACAGTGTAGCAATGACCTTACAGATGCTGACATTCGCGGCGGAGGTCGT 2258
Db 2185 CTGGGACAGAGCTGTACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2237
QY 2259 CAGTGGCAGACAGCCAGCAGAGATGATGATGATGATGATGATGATGATGATGAT 2297
Db 2238 CAGTGGCAGAGCTGTACAGCAGAGATGATGATGATGATGATGATGATGATGATGAT 2276

RESULT 9
AX685731
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS
BASE COUNT
ORIGIN
Query Match

AX685731 2019 bp DNA linear PAT 29-MAR-2003
Sequence 3 from Patent WO2081691.
AX685731
AX685731.1 GI:29371740
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Hobbs, H. H., Shan, B., Barnes, R. and Tian, H.
Abcg8 and abcg8: compositions and methods of use
Patent: WO 02081691-A 3 17-OCT-2002;
Tularik Inc. (US); BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)
Location/Qualifiers
1..2019
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
1..2019
/note="unlabeled protein product; mouse Abcg8 (mAbcg8)"
/codon_start=1
/protein_id="CAD86571.1"
/db_xref="GI:29371741"
/translation="MAEKTEKTLQVNGVILQDASGLDLSFSSPSDNLSTPTSGOS
NLEVRDLTYVDIASOVPMFQLOAFKTPMHSOSDCEGLIRLSKRVSGOMLA
ITGGSGGASLADVTYGRGKGKMSQITWINGPSTPQIVKCAVAHRQDOLPN
LTVRETLAETIAQMRPFTSQARQKREDAVLAELRLQCATRVGNTVYREVSGGER
RVSIGVOLNMBGILIDEPSTGDSFTAHNLVTLTSLAAGNLVLTSLQPSDI
FLVFLDLVLTMTGTPYVGAQOMVQFTSIGHPCPRYNADPFDVLTSLIDRSKER
EVAITYEKAOSIALAFLKLVKQGFDDPLAKAEKELMTSTHTVSTLTLOPTDCTAYELP
GMIROFSTLIRQINDRDLPTLILHSEKALMSLTIGFLTYGAGARQSLMDTAL
LPMIALIPENYILDVVSCHSEKSMLEYLEDGLTAGPYFAKILDELPERCAVYI
ISYAMPITVNTLRPVELFLAFLFLLVWLTVVCCRTMALASMLPTFPMSSFCNALLY
NSFYLTAGFMINDMLMIVPAWISKLSTLMFCFSGIMQIOFNGHLYTTOIGNFTSIL
GDTMSADLNHPLVYAIYLIYVIGSYGFLFVLYLSLKLINOKSIQDM"

444 a 598 c 510 g 467 t

53.6%; Score 1430; DB 6; Length 2019;

Best Local Similarity 82.0%; Pred. No. 5.3e-306;
Matches 1659; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

OY 100 ATGGCGCGGAAAGGCGGCAGAGAGAGAGAGGCGTCCGAAAGGGGCCACTGCCAGATAC 159
 Db 1 ATGGCTGAGAAAAACCAAGAAAGAACCCAGCTGTGGAAAGGAGACTGTACTTACAGATGCT 60
 OY 160 TCGGGCCATCCAGATAGATTGTTCTCCTCGTGAAGTGAACACAGGCTGTACTTCAACATC 219
 Db 61 TCGGGCCATCCAGAGACAGCTTGTTCCTCTCGGAAGTGAACACAGTCTGTACTTCAACATC 120
 OY 220 ACTGGCCAGGCCACACCCTGGAGGTCAAGAGACTCAATCTACAGTGGAGCTGGACCTCT 279
 Db 121 AGTGTGATAGTCCAAACACTCTGAGAGTCAAGAGATCTCACTACCAAGTGGACATGCCCTCT 180
 OY 280 CAGGTGCCCTTGGTTTGAGCAGCTGCTCAGTTCAAGATGCCGTGACATCTCCAGCTGC 339
 Db 181 CAGGTGCCCTTGGTTTGAGCAGCTGCTCAGTTCAAGATGCCGTGAGAGCTCATATGACAGC 240
 OY 340 CAGAAATTTCTGTGAGCTGGGACCTCAGAAACCTTAACCTTCAAGTGAAGTGGAGATG 399
 Db 241 CAGAGCTCTGTGAGCTGGGACCTCGAATCTTAACCTTCAAGTGAAGTGGAGACAGATG 300
 OY 400 CTGGCCATCATATAGGAGGCTCAGGTTGTGGAGAGGCTCTTGCTAGATGTGATCACTGAC 459
 Db 301 CTGGCCATCATATAGGAGGCTCAGGCTGGCGGAGAGGCTCACTACTGAGCTGATTCACAGGC 360
 OY 460 CGAGTCAACGGCGGCACAAATCAATGTCAGGCCAGATCTGATCAATGGGCAAGCCAGCTGC 519
 Db 361 AGAGGCCACGGGTGGCAGATGAAATCAAGGACAAATTTGGATTAATGGGCAACAGTACG 420
 OY 520 CCTCAGCTGTGAGGAGATGTGTGGCCCAAGCGGCCCAAGCAACACAGCTGTCCCCAAC 579
 Db 421 CCTCAGCTGTGAGGAGATGTGTGGCCCAAGCGGCCCAAGCAACAGCTGTCCCCAAC 480
 OY 580 TTGACTGTGCGAGAGACCTTGGCCCTTCAATGGCCAGATGCGGCTGCCAGAACTCTTCC 639
 Db 481 CTGACCGTCAGAGAGAACCTTGGCTTCAATGGCCAGATGCGGCTGCCAGAACTCTTCC 540
 OY 640 CAGGCCCAAGCTGACAAAAGGTGGAGAGACGTGATCGCGAGCTGCGGCTTAGCGATGC 699
 Db 541 CAGGCCCAAGCTGACAAAAGGTGGAGAGACGTGATCGCGAGCTGCGGCTTAGCGATGC 600
 OY 700 GCTGACACCCCGTGGGCGCAATGATGACCTGGGGGGTGTGCGGGGGGTGAGCGCGAGGA 759
 Db 601 GCGAACACCAAGTGGGCGCAACAGTATGATGAGTGGGGTGTGCGGGGGGTGAGCGCGAGGA 660
 OY 760 GTCAAGATTTGGGGGTGACGCTCTCTGTGAAACCCAGGAATCTTATTTCTGACAAACCCACC 819
 Db 661 GTGAGCAATTTGGGGGTGACGCTCTCTGTGAAACCCAGGAATCTTATTTCTGACAAACCCACC 720
 OY 820 TCTGGGCTCGACAGCTTTCACAGCCCAAAACCTGTGTGAAGACCTTGTCCAGGCTGGCCAAA 879
 Db 721 TCTGGGCTCGACAGCTTTCACAGCCCAAAATTTGTGTGAAGACCTTGTCCAGGCTGGCCAAA 780
 OY 880 GCGAACCGGCTGTGCTCATCTCCCTCCACAGGCTGCTGTGACATTTAGGCTGTTTT 939
 Db 781 GCGAACCGGCTGTGCTCATCTCCCTCCACAGGCTGCTGTGACATTTAGGCTGTTTT 840
 OY 940 GATCTGCTCTCTGATGATGCTGTGGCAACCCCATCTACTTAAGGGCGGCCACACATG 999
 Db 841 GACCTGTCTCTCTGATGATGATGCTGGCAACCCCATCTACTTAAGGGCGGCCACACCAATG 900
 OY 1000 GTTCAGATTTCTACAGCAGCTGGGTACCCCTGTCTCTGCTACACATCCCTGCTGACATTC 1059
 Db 901 GTTCAGATTTCTACATCTATGCTGGCACCCCTTGTCTCTGCTATTAACCAACCCCTGCGGACTTC 960
 OY 1060 TATGTGACCTGACACAGCATTTGACAGGCGCAGCAGAGAGCAGGAATTTGGCCACAGGAG 1119
 Db 961 TACGTGACTTTGACACAGCATTTGACAGGCGCAGGAAGAGAGGAGTGGCCACCTGTGGAG 1020
 OY 1120 AAGGCTCAGTACCTCGACGCCCTTCTTCTGAAAAAGTGGCTGACTTAATGACTTTCTTA 1179
 Db 1120 AAGGCTCAGTACCTCGACGCCCTTCTTCTGAAAAAGTGGCTGACTTAATGACTTTCTTA 1179

Db	1021	AAGGACAGTCTCTTGGACGCCCTGTCTCTAGAAAAAGTACAGGCTTTGATGACTTTCGTG	1080
Oy	1180	TGGAAGCAGAGAGCAAGAGATCTTGACAGAGGACACCTGTGTGAAAGCAGCGTACCCCA	1239
Db	1081	TGGAAGGTGAGCAAGGAACTCAACACACAGACCCACACACAGTACAGCTGACCTCACA	1140
Oy	1240	CTAGACACCAACTGCTCCGAGTCTTACGAAATGCGTGGGCGGTGACAGATTACG	1299
Db	1141	CAGGACACTGACTG---TGGGACGTCTGTTGAGCTGCGCCGGGATGATAGACAGTTTTC	1197
Oy	1300	ACCGGATCGGCTGCGCAGATTTCACAGACACTTCCGAAACCTCGGACCCACCCTCTATCAT	1359
Db	1198	ACCGTATCCCTGTGCAAGATTTCCAATGACTTCCGGGACCTGGACCGCTGCTATTCAT	1257
Oy	1360	GGGGCGAGGCGCTGTCTGATGTCAATGACCAATCGGCTCCCTCAATTTTGGCCATGGGAGC	1419
Db	1258	GGGTGCGAAGCCTGCTGATGATCCCTCATCTTGTGCTCTTACTACGGCAGTGGGGCC	1317
Oy	1420	ATCCAGCTCTCTTCATGATGATACAGCGGCCCTTGTTCATGATCGGTGCTCATCCCT	1479
Db	1318	AAGCAGCTCTCTTCATGATGACACAGACGCCCTCTCTTCATGATAGAGGGGCTATTCCT	1377
Oy	1480	TTCAACGTCATCTTCGATGTATCTTCCAAATGTTACTACAGAGGCAATGCTTACTAT	1539
Db	1378	TTCAATGTCATCCGGAATGTCTGCCAAATGTCTCCAGATGTCAGAGGTCAATGCTACTAT	1437
Oy	1540	GACGTGGAAGAGGGGCTGTACACACAGTGTCCATTTCTTGGCAAGATCCCTGGGAG	1599
Db	1438	GAGCTGGAGAGCGGGCTGTACACGTGTGTCTTATTTCTTGGCAAGATCCAGAGAA	1497
Oy	1600	CTTCCGAGGACCTGTGCTACATATCATATACAGGATGGCCACCTACTGGCTGGCCAC	1659
Db	1498	TTGGCGGAGGACCTGTGCTACATATCATATACAGGATGGCCACCTACTGGCTGGCCAC	1557
Oy	1660	CTGAGGCGAGGCTCTCCAGCCCTTCTGCTGCACTCTCTGTGTGTGTGTGTGTCTTC	1719
Db	1558	CTGGGCGCGGCTGTGAGCTCTCTCTCTCACTCTCTGCTCGATGTGTGTGTGTGTCTTC	1617
Oy	1720	TGTTGAGAGATATGAGGCGCTGGGCGGCGGCGGCGCTGTGCCACGCTTCCACATGGCTCC	1779
Db	1618	TGCTGCAAGACATGAGGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTTC	1677
Oy	1780	TTCTTCAGCAATGCGCTCTCACTACTCTTCTTACCTGCGCGGGGGCTCATGATAACTTG	1839
Db	1678	TTCTTCAGCAATGCGCTCTCACTACTCTTCTTACCTGCGCGGGGGCTCATGATAACTTG	1737
Oy	1840	AGCAGCCTGTGTGAGAGTGTGCGCGGCTGTGATTTCCAAAGTGTCTTCTGCGGTGTGT	1899
Db	1738	GACACACTGTGGAGTGTGCTGTGATGTATCCCAAGCTGTGCTTCCGCTGGTGTCTTC	1797
Oy	1900	GAAAGGCTGATGAGATTCAGTTACAGAGAAACTATTAATGGCTCTCGGGAACCTTC	1959
Db	1798	TGCGGGCTGATGAGATTCATTTAATGACACCTTTACACCAACAAATGGCAACTTC	1857
Oy	1960	ACCATCGCGGTCTCAGAGATTAATAATCTTCAAGTGCATGAGGTGATGCTGACCTCTTC	2019
Db	1858	ACCTTCTCANTCTCGGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT	1917
Oy	2020	TAGGCACTTCACTCATCTGCTGATTTGGGCTCTGACGGGTGCTTCAATGCTCTGATACGTG	2079
Db	1918	TATGCGGATCTCACTCATCTGCTGATTTGGGCTCTGATGATGATGATGATGATGATGATGAT	1977
Oy	2080	TGCTTGAAGTTCAATCAACAGAAACCAAGTCAAGACTGGGA 2121	
Db	1978	TGCTTGAAGTTCAATCAACAGAAACCAAGTCAAGACTGGGA 2019	

KEYWORDS
SOURCE
ORGANISM

HTG.

Homo sapiens (human)

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 127066)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone CTR-2367F13
Unpublished

TITLE
JOURNAL
AUTHORS

2 (bases 1 to 127066)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K., MCPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tittel, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
AUTHORS

Direct Submission

Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 127066)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., MCPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
AUTHORS

Direct Submission

Submitted (24-AUG-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 127066)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., MCPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Nguyen, C.,

TITLE
JOURNAL
COMMENT

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 11, 2001 this sequence version replaced g1:15284200.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

FEATURES

source

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L11578
Center clone name: 2367_F13

Location/Qualifiers

1. 127066
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="CTR-2367F13"
/clone.lib="CTB Human BAC"
complement(8..170)
/rpt_family="MER47A"
171..468
/rpt_family="AluSx"
complement(469..516)
/rpt_family="MER47A"
988..1049
/rpt_family="MIR"
complement(1294..1449)
/rpt_family="L1ME4A"
complement(2662..2954)
/rpt_family="AluSx"
4049..4431
/rpt_family="L2"
5261..5269
/note="<30 qual SNGL region"
7192..7202
/note="<30 qual SNGL region"
7310..7472
/rpt_family="MIR"
7488..7582
/rpt_family="MIR"
7589..7628
/rpt_family="MIR"
complement(7631..7781)
/rpt_family="AluSg/x"
7791..7922
/rpt_family="MIR"
complement(7977..8300)
/rpt_family="AluUb"
9044..9343
/rpt_family="AluSg"
10315..10344
/rpt_family="AT-rich"
10355..10681
/rpt_family="AluIo"
10683..10993
/rpt_family="AluSx"
complement(12221..12282)


```

repeat_region /rpl_family="MIR3"
complement(12306..12449)
/rpl_family="MIR"
complement(13008..13189)
/rpl_family="MER33"
complement(13190..13471)
/rpl_family="AluJo"
complement(13472..13612)
/rpl_family="MER33"
complement(13922..13899)
/rpl_family="GC-rich"
complement(14184..14250)
/rpl_family="L2"
complement(14630..14552)
/rpl_family="MER5A"
complement(14809..15100)
/rpl_family="AluSx"
complement(15363..15679)
/rpl_family="AluY"
complement(15681..15979)
/rpl_family="AluSx"
complement(16388..16292)
/rpl_family="L2"
complement(16508..16392)
/rpl_family="MUT1"
complement(16538..16616)
/rpl_family="LTR37B"
complement(16687..16618)
/rpl_family="Alu"
complement(16988..17104)
/rpl_family="L2"
complement(17895..17340)
/rpl_family="MUT1A1"
complement(17911..18209)
/rpl_family="AluS4"
complement(18680..18487)
/rpl_family="LTR16A1"
complement(19026..18802)
/rpl_family="AluJo"
complement(19092..19390)
/rpl_family="AluJo"
complement(21369..21675)
/rpl_family="AluSx"
complement(22474..22763)
/rpl_family="MER115"
complement(22843..22942)
/rpl_family="MER115"
complement(23311..23239)
/rpl_family="L3"
complement(23968..24265)
repeat_region

Query Match 27.1%; Score 724; DB 9; Length 127066;
Best Local Similarity 90.0%; Pred. No. 1,1e-149;
Matches 824; Conservative 0; Mismatches 5; Indels 87; Gaps 1;

```

```

Db 59738 CCATGAGCTGAGCTCGTACCTCTCTACAGCCATCTACCTCATTCGTCATTTGGCCTCAGCG 59737
Qy 2054 GTGGCTTCATGAGTCTCTGATCTACCTGTCCTTAAAGTTGATCAAGAAACAACTCAG 2113
Db 59738 GTGGCTTCATGAGTCTCTGATCTACCTGTCCTTAAAGTTGATCAAGAAACAACTCAG 59857
Qy 2114 ACTGCTGATTCACAGCCAGACGCTCTCCGCTGCTGGGAGCTTACAGACCTTCAACT 2173
Db 59858 ACTGCTGATTCACAGCCAGACGCTCTCCGCTGCTGGGAGCTTACAGACCTTCAACT 59917
Qy 2174 GCACCTCCCTCTCAGAGACCCCTTCTTGGGACAGTACAGACATGACCTTACAGATCT 2233
Db 59918 GCACCTCCCTCTCAGAGACCCCTTCTTGGGACAGTACAGACATGACCTTACAGATCT 59977
Qy 2234 CAGCTACATCCGGCCAGGCTGCTCAGTGGACAGACAGACAGACAGATGGCATAGAA 2293
Db 59978 CAGCTACATCCGGCCAGGCTGCTCAGTGGACAGACAGACAGACAGATGGCATAGAA 60037
Qy 2294 TAAAGACAGTGAAGGATTTCTGCTCAGTGGACAGAGACTGCTATGATGATGAT 2353
Db 60038 TAAAGACAGTGAAGGATTTCTGCTCAGTGGACAGAGACTGCTATGATGATGAT 60097
Qy 2354 CTGCTACCTGGTGGCAGCTTACAGCTGCTATATTTCTTCTTGTATGATGAT 2413
Db 60098 CTGCTACCTGGTGGCAGCTTACAGCTGCTATATTTCTTCTTGTATGATGAT 60157
Qy 2414 AGGCACTGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2473
Db 60158 AGGCACTGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60217
Qy 2474 GAATTTGTTGAACCTGAGAGGACATTAACATGATGATGATGATGATGATGATGAT 2533
Db 60218 GAATTTGTTGAACCTGAGAGGACATTAACATGATGATGATGATGATGATGAT 60277
Qy 2534 GGGGCCCCACACTCCGCTGCTGAGCCACCATCATACAGAAAGTACCTTACAG 2593
Db 60278 GGGGCCCCACACTCCGCTGCTGAGCCACCATCATACAGAAAGTACCTTACAG 60337
Qy 2594 CAGATGCTATCCCTCTTCTTGTGGGCTGATGCTCCAAAGCAGTACAT 2653
Db 60338 CAGATGCTATCCCTCTTCTTGTGGGCTGATGCTCCAAAGCAGTACAT 60397
Qy 2654 TAAAAATGATTTAGC 2669
Db 60398 TAAAAATGATTTAGC 60413

RESULT 11
AC087053 182261 bp DNA linear HTG 27-MAR-2003
DEFINITION Homo sapiens chromosome 2 clone RP11-959M3 map 2, 9 unordered
ACCESSION AC087053
VERSION AC087053.13 GI:25140148
KEYWORDS HTG; HTGS-PHASE1; HTGS-FULLTOP; HTGS-CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 182261)
AUTHORS Birren,B., Nussbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-959M3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182261)
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Bouhagalter,B., Brown,A.,
Cammarata,J., Campoliano,A., Choquet,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Ferreira,P., FitzHugh,M., Gage,D., Galagan,J.,
Gardyna,S., Glende,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Lacroque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,

```


OY 2594 CAGATGCCATCCCTTTTGTGCGGGTCATGGGCTCCAAAGCCACAGTCAGCAAT 2653
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 7710 CAGATGCCATCCCTTTTGTGCGGGTCATGGGCTCCAAAGCCACAGTCAGCAAT 7769
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 2654 TAAAAATGATTGAGC 2669
 ||||||||||||||||||
 DB 7770 TAAAAATGATTGAGC 7785
 ||||||||||||||||||

RESULT 12
 AC108476 139342 bp DNA linear PRI 16-APR-2002
 LOCUS AC108476
 DEFINITION Homo sapiens BAC clone RP11-1413K20 from 2, complete sequence.
 AC108476
 VERSION AC108476.5 GI:19807988
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo;
 1 (bases 1 to 139342)
 Sultson, J.E. and Waterston, R.
 JOURNAL Toward a complete human genome sequence
 GENOME 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 139342)
 AUTHORS Hartings, C., Haekenson, W. and Doeber, A.
 TITLE The sequence of Homo sapiens BAC clone RP11-1413K20
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (27-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 139342)
 REFERENCE 4 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 139342)
 REFERENCE 5 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 6 (bases 1 to 139342)
 REFERENCE 6 (bases 1 to 139342)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Mar 29, 2002 this sequence version replaced gi:18767626.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH1413K20

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
 Tateo, M., Cattanese, J.J., and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at <http://www.chori.org>
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-489K22, 2000 bp overlap.
 Actual end is at base position 139342 of RP11-1413K20.

The region between 132012 and 132017 is covered only by a PCR
 product of clone DNA.

FEATURES

source
 1..139342
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-1413K20"
 /clone_11b="RPCT-11"
 55..653
 /note="match to EST AA203458 (NID:g1799169) zx58b04.r1"
 misc_feature
 93..286
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 93..286
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 vj29c06.y1"
 93..279
 /note="match to EST AV660973 (NID:g9881987)"
 misc_feature
 318..653
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 misc_feature
 372..633
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 misc_feature
 ye58h05.r1"
 706..708
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 misc_feature
 706..707
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 repeat_region
 847..1139
 /rpt_family="Alu"
 misc_feature
 1867..2047
 /note="match to EST T39945 (NID:g647612) ya13g04.r1"
 repeat_region
 2234..2616
 /rpt_family="L2"
 misc_feature
 2983..3121
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 vj29c06.y1"
 3044..3121
 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
 misc_feature
 4099..4304
 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
 misc_feature
 4099..4283
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 4401..4618
 /note="similar to Mus musculus EST BF162656

1841 GCAGCGTGTGGACAGTGCCTGGATTTCACAAAGTGCCTTCGCGGTGTGTTTG 1900

RESULT	13
F351812S13	
LOCUS	F351812S13
DEFINITION	Homo sapiens sterolin-2 (ABCG8) gene, exon 13 and complete cds.
ACCESSION	AF351824
VERSION	AF351824.1 GI:15146443
KEYWORDS	
SEGMENT	13 of 13
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H., Ose, L., Stalenhoef, A.F., Miettinen, T., Bjorkhem, I., Bruckert, E., Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and Patel, S.B.

TITLE

Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterol-1 and sterol-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL
MEDLINE
PUBMED

Am. J. Hum. Genet. 69 (2), 278-290 (2001)

REFERENCE
AUTHORS

Lu, K.
Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

FEATURES

Location/Qualifiers
1..2201

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="between D2S2294 and D2S2298"

/clone="108162; 32814"

/cell_type="ES cell"

/order="AF351812.1:1227..2809, AF351813.1:1..4665, AF351814.1:1..2368, AF351815.1:1..1323, AF351816.1:1..300, AF351817.1:1..660, AF351818.1:1..203, AF351819.1:1..888, AF351820.1:1..685, AF351821.1:1..884, AF351822.1:1..1292, AF351823.1:1..859, 1..>182)

/gene="ABCG8"

/product="sterolin-2"

/protein_id="AAK84663.1"

/translation="MAGKAAFEKRLPKGATPDNSGLDRIPLFSSSDNSLYPTSGOP

ITIGSGGCRASLIVITGKGHGKIKSQIINQSPSPOLVRCCVAVRHNOLPLN

LRVETLAFIAOMRLPTFSOARDKREVDIAELRPLQDRTVGNVAVRSGGER

FRVFDVILMTNPGILIDEPISGLDSTANLVKTSLAKNRLVLSIHSORDI

ELATREKQSLIALELEKVRIDDEPKAEKRLDEPTCEVSTYPLDNCPSPTKAP

GAVOQFKTLIRQISNDPRDPTLILHACACLMKMTGFLYFGHSIQLSFMDYAL

LPMLGALIPNVILIDVSKCYSERAMLYELEDELDTYTPPEFAKIIIGEPHAYII

ITGMPTWMLANRGLQPLFLHPLVLMVYCCIMLAAALPTFIMASFFSNALY

NSFIAGGFMINLSLMTVPAMISVFLKRCFGLMKIORSRRYKPKGLNLTAVS

GDKILSMEIDSYPLVAYILVIGLSGGFWLYVSLRFIKRQSDM"

45..>182

/gene="ABCG8"

/number=13

BASE COUNT

543 a 536 c 538 g 580 t 4 others

ORIGIN

Query Match

24.9%; Score 663.6; DB 9; Length 2201;

Best Local Similarity 99.3%; Pred. No. 3,4e-136;
Matches 685; Conservative 2; Mismatches 1; Indels 2; Gaps 2;

QY 1982 AATCTCAGTGCATGAGAGAGAGTGTGACCTCTCTACAGCCATCTACCCATCAGTCA 2041

DB 43 AGATCTCAGTGCATGAGAGAGTGTGACCTCTCTACAGCCATCTACCCATCAGTCA 102

QY 2042 TTGGCTCAGCGGTGAGCTTCATGCTCTCTACAGTGTCTTAAGGTTCATCAACAGA 2101

DB 103 TTGGCTCAGCGGTGAGCTTCATGCTCTCTACAGTGTCTTAAGGTTCATCAACAGA 162

QY 2102 AACCAAGTAAAGCTGTGATTCACGCCAGACGTCTGCCGTGTGGGGACCTGAGCA 2161

DB 163 AACCAAGTAAAGCTGTGATTCACGCCAGACGTCTGCCGTGTGGGGACCTGAGCA 222

QY 2162 GACCCCTCAAGTCACTCCCTCTCAGAGGCCCTTCCTGGGGACAGTACATATAC 2221

DB 223 GACCCCTCAAGTCACTCCCTCTCAGAGGCCCTTCCTGGGGACAGTACATATAC 282

QY 2222 CCTACAGATGCTACGATACATCCGGCCAGGTGTGTCAGTGCACAGACAGCCAGAG 2281

DB 283 CCTACAGATGCTACGATACATCCGGCCAGGTGTGTCAGTGCACAGACAGCCAGAG 342

QY 2282 ATGGCAGTGAATTAAGACAGTCAAGAGGATTTCTCTACCTGACAGAGACTGCATG 2341

DB 343 ATGGCAGTGAATTAAGACAGTCAAGAGGATTTCTCTACCTGACAGAGACTGCATG 402

QY 2342 ACTGGGAGAAACCTGCACCTCCGCTGACCTCAACAGCTTCTTAATTTCTTTGAT 2401

DB 403 ACTGGGAGAAACCTGCACCTCCGCTGACCTCAACAGCTTCTTAATTTCTTTGAT 462

QY 2402 ATGCATTTATATAGCACTCGATATAGATGGAGCAACAGTGAATTTGGTGGC 2461

DB 463 ATGCATTTATATAGCACTCGATATAGATGGAGCAACAGTGAATTTGGTGGC 522

QY 2462 TAGACTGTGACGAATTTGTGGACCTGGAGGAAATTAACAGTACAGTACAGATTGG 2521

DB 523 TAGACTGTGACGAATTTGTGGACCTGGAGGAAATTAACAGTACAGTACAGATTGG 582

QY 2522 CTTCACTTCCAGGGGGCCCACTCCCTGTGTGAGCCACCATCAATACAGAAAGTACCT 2581

DB 583 CTTCACTTCCAGGGGGCCCACTCCCTGTGTGAGCCACCATCAATACAGAAAGTACCT 642

QY 2582 -AAGATGTACGACGACGATGATCCATCTCTTTTGTGTGGGGTCCGCTCCAAAGC 2640

DB 643 -AAGATGTACGACGACGATGATCCATCTCTTTTGTGTGGGGTCCGCTCCAAAGC 702

QY 2641 CAACGTGAACAA-TTAAAAATGTATTTGAGC 2669

DB 703 CAACGTGAACAA-TTAAAAATGTATTTGAGC 732

RESULT 14

F351812506

LOCUS F351812506 660 bp. DNA linear PRI 10-AUG-2001

DEFINITION Homo sapiens sterol-2 (ABCG8) gene, exon 6.

ACCESSION AF351817

VERSION AF351817.1 GI:15146436

KEYWORDS 6 of 13

SEGMENT Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H., Ose, L., Stalenhoef, A.F., Miettinen, T., Bjorkhem, I., Bruckert, E., Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and Patel, S.B.

Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterol-1 and sterol-2, encoded by ABCG5 and ABCG8, respectively

Am. J. Hum. Genet. 69 (2), 278-290 (2001)

MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 660)
 AUTHORS Lu, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St., STB 541, Charleston, SC 29403, USA

FEATURES
 source location/Qualifiers
 1. 660
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="between D2S2294 and D2S2298"
 /clone="1081G2, 32814"
 /cell_type="ES cell"
 66. 335
 /gene="ABCG8"
 /number=6

exon
 BASE COUNT 158 a 184 c 172 g 146 t
 ORIGIN

Query Match 10.3%; Score 274.2; DB 9; Length 660;
 Best Local Similarity 97.2%; Pred. No. 6.8e-50;
 Matches 279; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 781 CTGTGGAACCCAGGATTCCTTATTCGAGAACCCACCTCTGGCTCCACAGCTTCACA 840
 |||||
 DB 53 CTGTGTGGAAGAAATCCCTTATTCGAGAACCCACCTCTGGCTCCACAGCTTCACA 112
 |||||

QY 841 GCCCAACCTGTGGAAGACCTGTCCAGGCTGGCCAAAGGCAACGGGGTGGTCATC 900
 |||||
 DB 113 GCCCAACCTGTGGAAGACCTGTCCAGGCTGGCCAAAGGCAACGGGGTGGTCATC 172
 |||||

QY 901 TCCCTCCACAGGCTGTCTGACATCTTCAGGCTTTTGATCTCTCTCTGATGACG 960
 |||||
 DB 173 TCCCTCCACAGGCTGTCTGACATCTTCAGGCTTTTGATCTCTCTCTGATGACG 232
 |||||

QY 961 TCTGGGACCCCATCTTACTTAGGGGGGGCCACACATGATGTCAGTATTTACAGCCATC 1020
 |||||
 DB 233 TCTGGGACCCCATCTTACTTAGGGGGGGCCACACATGATGTCAGTATTTACAGCCATC 292
 |||||

QY 1021 GGCTACCCCTGTCTGCTACAGCAATCCGCTGACTTCATGTGA 1067
 |||||
 DB 293 GGCTACCCCTGTCTGCTACAGCAATCCGCTGACTTCATGTGA 339
 |||||

RESULT 15
 F351812S11 1292 bp DNA linear PRI 10-AUG-2001
 LOCUS F351812S11
 DEFINITION Homo sapiens sterol-in-2 (ABCG8) gene, exon 11.
 ACCESSION AF351822
 VERSION AF351822.1 GI:15146441
 KEYWORDS
 SEGMENT 11 of 13
 SOURCE Homo sapiens
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1292)
 AUTHORS Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
 Ose, L., Stalenhoef, A.F., Miettinen, T., Bjorkhem, I., Bruckert, E.,
 Pandya, A., Brewer, H.B., Jr., Salen, G., Dean, M., Sirtavastava, A. and
 Patel, S.B.
 TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterol-in-1 and
 sterol-in-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
 MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 1292)
 AUTHORS Lu, K.

TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St., STB 541, Charleston, SC 29403, USA

FEATURES
 source location/Qualifiers
 1. 1292
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="between D2S2294 and D2S2298"
 /clone="1081G2, 32814"
 /cell_type="ES cell"
 101. 368
 /gene="ABCG8"
 /number=11

exon
 BASE COUNT 233 a 405 c 317 g 327 t 10 others
 ORIGIN

Query Match 10.2%; Score 273.4; DB 9; Length 1292;
 Best Local Similarity 86.7%; Pred. No. 9.9e-50;
 Matches 301; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1586 AGATCCCTGGGGAGCTTCGGAGCAGCTGCTCAATCATCATACGAGATGCCACCT 1645
 |||||
 DB 99 AGATCCCTGGGGAGCTTCGGAGCAGCTGCTCAATCATCATACGAGATGCCACCT 158
 |||||

QY 1646 ACTGGCTGGCCAACTGAGGCCAGGCTCCAGCCCTTCGCTGCACTTCCTGGTGT 1705
 |||||
 DB 159 ACTGGCTGGCCAACTGAGGCCAGGCTCCAGCCCTTCGCTGCACTTCCTGGTGT 218
 |||||

QY 1706 GCGTGTGCTCTTCTGTGACAGATTATGAGCCCTGCGCGCGGCTGCTCCACCT 1765
 |||||
 DB 219 GCGTGTGCTCTTCTGTGACAGATTATGAGCCCTGCGCGCGGCTGCTCCACCT 278
 |||||

QY 1766 TCCACATGGGCTCTCTTCCAGCAATGCCCTCTACAACTCTTACCTGCGGGGGCT 1825
 |||||
 DB 279 TCCACATGGGCTCTCTTCCAGCAATGCCCTCTACAACTCTTACCTGCGGGGGCT 338
 |||||

QY 1826 TCATGATTAACCTTGAGCAGGCTGTGACAGTGGCCCGTGATTTCCAAAGTGTCTCC 1885
 |||||
 DB 339 TCATGATTAACCTTGAGCAGGCTGTGACAGTGGCCCGTGATTTCCAAAGTGTCTCC 398
 |||||

QY 1886 TCGGCTGTGCTTTTGAAGGCTGATGAAGATTCAATTCAGCAGAGA 1932
 |||||
 DB 399 TCGGCTGTGCTTTTGAAGGCTGATGAAGATTCAATTCAGCAGAGA 445
 |||||

Search completed: July 27, 2003, 05:52:39
 Job time : 9494.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 23:50:29 ; Search time 4963.41 seconds
(without alignments)
13069.363 Million cell updates/sec

Title: US-09-989-981a-7
Perfect score: 2669
Sequence: 1 gtgtccctgtccaggaac.....caatlaaataatgatgac 2669

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est4:*
13: gb_est3:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1511.6	56.6	3623	11 AK004871	AK004871 Mus muscu
2	1286.2	48.2	2417	11 AK050938	AK050938 Mus muscu
3	681.8	25.5	691	2 HSM072006	Bx481838 Homo sapi
4	460.6	17.3	849	12 B1330745	B1330745 602982409

5	370.4	13.9	549	10 B660076	B660076 maa27c08.
6	361.4	13.5	583	14 BY705076	BY705076 BY705076
7	355.8	13.3	457	14 T91380	T91380 yd53b02.s1
8	332.4	12.5	334	2 HSM072530	Bx483362 Homo sapi
9	331.2	12.4	510	10 HSM072530	B6610072
10	323	12.1	511	9 A157365	A157365 u45h01.y
11	318.4	11.9	564	14 T84531	T84531 u45b02.r1
12	309.8	11.6	500	9 A115181	A115181 u46c10.y
13	276.8	10.4	463	9 AA537862	AA537862 v35a03.r
14	276	10.3	781	14 CB502603	CB502603 ssalimgc50
15	234.6	8.8	398	9 A1597406	A1597406 v35a03.y
16	226.6	8.5	586	11 AK008188	AK008188 Mus muscu
17	226.4	8.5	581	14 BY708144	BY708144 BY708144
18	202.8	7.6	435	13 BX099922	BX099922 BX099922
19	198.4	7.4	435	9 A1574075	A1574075 u167h11.y
20	196.4	5.6	821	29 B2650554	B2650554 OGCBA897C
21	149.2	5.2	523	6 AU195806	AU195806 Porphyra
22	145.4	5.4	447	12 B1145065	B1145065 602909138
23	144.2	5.4	936	10 BF162856	BF162856 601769307
24	141.6	5.3	823	14 CB649273	CB649273 OSJNEB13B
25	141	5.3	833	10 BF620684	BF620684 HVSMEC002
26	138.8	5.2	566	6 AU192726	AU192726 Porphyra
27	138	5.2	754	14 CB627408	CB627408 OS1EB02F
28	137.8	5.2	791	13 BO743924	BO743924 WHEA109.G
29	137	5.1	854	28 BH703879	BH703879 BOHUR247F
30	135.8	5.1	793	14 CB620418	CB620418 OS1EA05J
31	135	5.0	397	14 CA024091	CA024091 HZ48P10F
32	134.4	4.9	581	12 BJ245824	BJ245824 BJ245824
33	132	4.9	798	29 B2531631	B2531631 OGAFH497C
34	132	4.9	803	29 B2986050	B2986050 PUGKFC267C
35	132	4.9	822	29 B2731642	B2731642 OGCRCY267C
36	132	4.9	858	29 B2730374	B2730374 OGBE0437M
37	132	4.9	746	14 CB627374	CB627374 OS1EB02E
38	131.4	4.7	816	10 BG299523	BG299523 HVSMEC002
39	126.6	4.7	274	14 R89160	R89160 yq02a04.r1
40	126.4	4.7	490	29 B2614705	B2614705 1946b01.g
41	125.8	4.7	867	14 CD382895	CD382895 PTM08065
42	124.4	4.6	841	14 CB683273	CB683273 OSJNEF11H
43	123.6	4.6	563	9 AV835407	AV835407 AV835407
44	123.2	4.6	812	29 B2537267	B2537267 OGCAG2867C
45	122.6	4.6			

ALIGNMENTS

RESULT 1
LOCUS AK004871 3623 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300003C16 Product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full insert sequence.
ACCESSION AK004871 GI:12836380
VERSION AK004871.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE 20499374

PUBLISHED 11042159
REFERENCE

TITLE
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, Y., Nishi, K., Kitsuma, T., Tashiro, H., Itoh, M.,
Sun, L., Ishii, Y., Nakamura, S., Hazama, N., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiyagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawaji, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system—384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE 20530913
PUBLISHED 11076861
REFERENCE

TITLE
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Ikawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,
Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Cassavani, T.,
Fleischmann, M., Gaasterland, T., Gissi, C., Kling, B., Kochiya, H.,
Knehl, P., Lewis, S., Matsuo, Y., Nikolaic, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Stabili, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Anno, H.,
Baldarelli, R., Barsh, G., Blake, J., Boftelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Butt, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamlay, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Togo-Oka, K., Wang, K.H., Weltz, C., Whitaker, C., Wilming, L.,
Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuki, S.,
and Hayashizaki, Y.

JOURNAL
MEDLINE 21085860
PUBLISHED 11217851
REFERENCE

TITLE
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3623)

JOURNAL
REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Anno, H.,
Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoaka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Koijima, Y., Konno, H., Kouda, M.,
Koya, S., Kuhihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Saeki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGACGAGACGGCGCCCACTGCAGATTTTTTTTTTTTNN 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence 5' GAGGAGACAGACATCCAGAGTCATTTATTTATTAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES	source
Location/Qualifiers	1. 3623
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C57BL/6J"	
/db_xref="FANTOM.DB:1300003C16"	
/db_xref="MGI:1896857"	
/db_xref="taxon:10090"	
/clone="1300003C16"	
/sex="male"	
/tissue_type="liver"	
/clone_idb="RIKEN full-length enriched mouse cDNA library 69_2090	
/dev_stage="adult"	
/note="unnamed protein product; ATP-BINDING CASSETTE, SUB-FAMILY 6, MEMBER 8 (STEROLIN-2) homolog [Mus musculus (SWISSPROT Q9DEMO, evidence: FASTA, 928id, 96.7%length, match=1796)]	
putative"	
/codon_start=1	
/protein_id="BAB23630.1"	
/db_xref="GI:12836381"	
/db_xref="MGI:1914720"	
/translation="MAEKTEETLOMNGTVLADASOGLODSLSESSDNLSTFYSGSNTLEVRDLTYOVDAISQVPMPELOAKQRIPIRSHSNDSCSLGIRNLSFKRSGSALISGSSGSRASILDVITGRGRGGRKMGKQIVDPSTPOLYRRCVAHYRHDDLLNLYRETAFLAQNRILPRTFSOQRDKRVEDYIAELRLKQCNANTVGNATYVAGSGGRRRSIGVQLMNPGLILDEPTSGLSDFETANLVTTLSRLAKRNLVLSIADPRRSIFRLEVDILMTISGPYILGAQOQVQFTSGHPCPSYSPNADYVDLTSIDRKSIREVAVERAKOSLALFELEYGQDFDLMEAKREKELNTSHYVSLITLODGTAVEIPGMVEFTSLIRROISDNFRDPLTLLHGSEACMLSLIGFLYGHGAKOLSPMDTALLPMGALIPNPVLDVYSKCHSERSMITYELEDGLTAGPREFAKILGELPEHCATYIITYAMPITWLNLRVPPEFLFLHFLVNLVVFCCRTMLAASAMLPETHMSSEFCNALYNSFTFLWPNIDNLMIVPAMISKLSFLRCEFGLMQIOFNGLYTOIGNFTSGSGDMISAMDNLNHPHLYAIVLIVIGISYGFLEFLYLSLTIKOKSIODM"	
polyA_signal	
polyA_site	
BASE COUNT	847 a 1012 c 920 g 844 t
ORIGIN	
Query Match	56.6%; Score 1511.6; DB 11; Length 3623:
Best Local Similarity	77.0%; Pred. No. 0:
Matches 1965; Conservative	0; Mismatches 534; Indels 53; Gaps 8
OY	99 CATGGCCGGGANGGCGGAGAGAGAGAGAGGCGTGCAGAAAGGCGCCACTGCCAGAGATAC 158
DB	68 CATGGCTAGAGAAACCAAGAAAGAGACCAGCTGTGAGATGGAGACTGTACTTCAGAGATGC 127
OY	159 CTC--GGGCGTCCAGGATAGATTGTCTCTCTCTGAAGTAGCAACACCTGTACTTAC 215
DB	128 TTCGAGAGGCTCCAGAGACGCTGTCTCTCGGAAAGTAGCAACAGCTGTACTTCC 187
OY	216 CTACAGTGGCAGCCCAACACCTGTGAGAGTCAAGAGACCTCAACTACAGTGGAGACTGAC 275
DB	188 CTACAGTGTGATCAACACTCTGAGAGTCAAGATCTCAACCTACAGAGTGGAGATGCG 247
OY	276 CTCCTCAGGTCCTTGTGAGAGAGCTGGCTCACTTCAAGATGCCCTGTGACATCTCCAG 335
DB	248 CTCCTCAGGTCCTTGTGAGAGAGCTGGCTCACTTCAAGATGACCTGTGAGAGTCTCAATAG 307
OY	336 CTCGACAGATCTTGTGAGCTGGGATCCAGAACTAGCTTCAAGAGAGAGAGAGGAGCA 395
DB	308 CAGCCAAAGACTCTCTGTAGCTGGGCACTCCGAATATCTTAAGCTTCAAGATGAGAGGTGACA 367
OY	396 GATCGGCCCAATAGAGAGCTCAGGTTGTGGGAGAGCTCTTGTGATGTGATCAG 455

Db	368	GATGCTGGCCATCATAGGAGGCTCAGGCTCGGGAGAGCTCACTACTCGACGGATCAC	427
OY	456	TGGCCGAGGTCACGGCGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGCAGGCCAG	515
Db	428	AGCCAGAGGCGCACGGGTGGCAAGATGAATAATCAGACAAATTTGGATTAATGGCAACCCAG	487
OY	516	CTGGCCGTCAGCTGGGTAGAGAAAGTGTGGGCCACGTGCGCCAGACACACAGCTGCTCCC	575
Db	488	TACGGCTCAGCTGGTATAGAGAAAGTGGCTTGGCGCATGTGCGGCACATGACCACCACTGGCC	547
OY	576	CAACTTACTGTCGAGAGAGACTTGGCCCTCATTTGCCAGATGCGGTGGCCGCAAGACTT	635
Db	548	CAACCTGACCGTTCAGAGAGACCTTGGCTTCAATTTGCCAGATGCGGTGGCCGCAAGACTT	607
OY	636	CTCCAGAGGCCAGCTGACAAAAGGGTGAAGAGCTGATGCGCGAGCTGCGGCTTAGCA	695
Db	608	CTCCAGAGGCCAGCTGACAAAAGGGTGAAGAGCTGATGCGCGAGCTGCGGCTTAGCA	667
OY	696	GTCGGCTGACACCCCGGTGGGCAACTATACGTGCGGGGGTGTGCGGGGGTGAAGCGAG	755
Db	668	GTCGGCTGACACCGAGAGTGGCAACATGATATGACGTGGGGTGTGCGGGGGTGAAGCGCG	727
OY	756	GAGATGACAGATTGGGGGTGACGCTCGTGTGGAACCCAGAAATTCCTATTCTGACGACAC	815
Db	728	ACGATGAGACATTGGGGGTGACGCTCGTGTGGAACCCAGAAATTCCTATTCTGATTAACC	787
OY	816	CACCTCTGGGCTTCAGACTTCACAGCCCAACAACCTGATGAAGACTTGTCCAGGGTGGC	875
Db	788	CACCTCTGGGCTTCAGACTTCACAGCCCAACAATCTGTGTGACACACTTGTCCGCTGGC	847
OY	876	CAAAAGCAACCGGCTGTGCTATCTCCCTCCACAGCTCGCTGTGACATCTTACAGCT	935
Db	848	CAAGGCAACAGGCTGTGCTATCTCCCTCCACAGCTCGCTGTGACATCTTACAGCT	907
OY	936	GTTTGTATCGGCTCTGTATGACGCTGTGGCAACCCCATCTACTTAGGGGCGGCCAGCA	995
Db	908	ATTTGAACCTGTGCTCTGTATGACATCTGGCAACCCCATCTACTTAGGGGCGGCCAGCA	967
OY	996	CATGCTCAGTATTTTCACAGCCAGGGCTTACCCCTGTCTCGTACAGCAATCTGCTGA	1055
Db	968	AATGGGAGATCTTCATCATTCATTGGCCACCCCTGTCTCGTATAGCAACCTGTGCGGA	1027
OY	1056	CTTCTATGTGGACTGTACAGCATTTGACAGGGCGACAGAGACAGAAATTTGGCCACAG	1115
Db	1028	CTTCTATGTGGACTGTACAGCATTTGACAGGGCGACAGAGACAGAAATTTGGCCACAG	1087
OY	1116	GGAGAAGGCTCAGTCTCGACCTGCGAGCCCTGTTCTAGAAAAAGTGTGCTGACTTAGACTT	1177
Db	1088	GGAGAAGGCTCAGTCTCGACCTGCGAGCCCTGTTCTAGAAAAAGTGTGCTGACTTAGACTT	1147
OY	1176	TCTATGGAAGACAGACAGACGATCTTGACAGGACACCTGTGTGGAAGAGGAGGTGAC	1233
Db	1148	TCTGTGGAAGACGCTGTGACAGGACACCTGTGTGGAAGAGGAGGTGAC	1207
OY	1236	CCCACTAGACACCAATGCTCTCCAGAGTCTTACGAAGATGCTTGGGGGGTGCACAGATT	1295
Db	1208	CACACAGACACTGTACTGCTT--TGGAGTGTGTTAGAGTGGCCGGGAGTATGACACACTT	1266
OY	1296	TACGACGCTGATCCGCTGTCTGATCTTCCAGATTTCCAGACACTTCCGAGACTTCCACCTGAT	1355
Db	1265	TTTCACCCCTGATCCGCTGTCTGATCTTCCAGATTTCCAGACACTTCCGAGACTTCCACCTGAT	1322
OY	1356	CCATGGGGCGAGGCTGTGATGTTCATATGACCAATGGCTCCCTATTTTGGCCATGG	1415
Db	1325	TCATGGGTGCGGAAGCTGCTGTATGTCCCTATCATTTGGCTTCCCTTACTACGGGCAATGG	1384
OY	1416	GAGCATTCAGCTCTTCCTTCATGATGATACAGCGCCCTCTTGTTCATGATGCGGTGCTCAT	1475
Db	1385	GGCCAAAGAGCTCTTCCTTCATGATGATGATACAGCGCCCTCTTCCTTCATGATGAGGCGCTCAT	1444
OY	1476	CCCTTCAAGCTCATTTGATGTGATCTTCCAAATGTTACTGACAGAGCGCAATGCTTTA	1535

Db	1445	TCCTTCANATGTCATCTGGATATGTGCTCTCCCAAATATGTCACTGCGAGAGGTCAATCTCTGA	1504
QY	1536	CTATGACATGGAAGACGGGCTGTACACCACTGATGCCAATATTTCTTTGGCAAGATCTCTCG	1595
Db	1505	CTATGATGATGAAGACGGGCTGTACACTCTGCTGCTCTTATTTCTTTGGCAAGATCTCTAG	1566
QY	1596	GGAGCTTCCGGAGACATGTGCTTACATCATATCTACGGGATGCCACCTACTGCTGGC	1655
Db	1565	AGATATGCCGGAGACATGTGCTTACGTCACTATCAAGGATGCCACCTACTGCTGTAC	1624
QY	1656	CAACCTGAGGCCAGGCGCTCCACCCCTTCCGTGTGACCTTCCGCGGGGTGGCTGGTGT	1715
Db	1625	AAACCTGGGGCCCTGCTGCTGAGACTCTTCTTACACTTCCGCTGCTGGTGGTGGTGT	1684
QY	1716	CTTCTGTTGACAGAAATTTATGGCCCTGAGCCGCGCGCCCTGCTCCACCTTCCACATATGC	1775
Db	1685	CTTCTGCTGACAGACCATATGGCCCTGAGTGCCTGCTCANTGGTGGCCACCTTCCACATATC	1744
QY	1776	CTCTCTTCTTACAGCAATGCGCTTACAACTCTTCTTCACTCCGCGGGGCTTCATGATATA	1835
Db	1745	CTCTCTTCTTCAATATGCGCTTCAACCTCTTCTTCACTCTTCTGCGGCTTCATGATATA	1804
QY	1836	CTTTGACACGCTGTGTGACAGCCCGCGGGGATTTCCAAAGTGTCTTCCCTGCGGCTGTG	1895
Db	1805	CTTTGACACACCTGTGTGATATGTGCTGTGACATGGATTTCCAAAGTGTCTTCCCGGTGTG	1864
QY	1896	TTTTGAAGGGCTGATGAAGATTTCACTTACAGCAGAAGAACTATTAATAATGCTCTCGGGAA	1955
Db	1865	CTTTCGCGGGCTGATGACGATTTCAATTTAATGACACCTTTACACACACAATTCGGCAA	1924
QY	1956	CCTCACACATGCGGGTCTGAGAGATATAATCCTCAGTGCATATGAGCTGACACTGTATCC	2015
Db	1925	CTTTCACCTTCTTCCATCTCTCGGAGACACATGATATAGTGCATATGAGACCTGAACTGTATCC	1984
QY	2016	TTCTTACAGGCATCTTCACTCATGTCATGTCATTTGGCGCTCAGAGGTTGGCTTCAATGCTCTGTACTA	2075
Db	1985	ACTGTATGCGATCTACCTCATTTGTCATGCGATAGCTATACGGCTCTCTTCTCTGTACTA	2044
QY	2076	CGTGTCTTAAAGTTTCATCAAAACAGAAACCAAGTCAAGACAGTGTGATTTACAGCCACAGCT	2135
Db	2045	TTCTATCTTGAAGCTCATCAAAACAGAACTCAATTCAGACTGTGATTACTCAGCTTGTCT	2104
QY	2136	CTGCCCCGTGGTGGGGGACCTGACAGACCTTCACTGACACTCCCTCTCAGAGAGCCCC	2195
Db	2105	CTCATCTGCGCG-----GACCCCTTTTCCCGGGGGCTGGCCACCCACAGAGAGAGCC	2152
QY	2196	TTTCCGGGGGACAGTGAAGACAATATACCTTACAGATGCTCAAGTACATCCGAGCCAGGGTG	2255
Db	2153	GGAGTGGGGGAAAGGCTCTCACACATATCTTCAG-----GCACACACCCACCTTCTTGTGTG	2205
QY	2256	CTGCACTGGGACAGACCCAGCCAGAGAGATGGAGTAAAGCAATCCGAAAGGATTT	2315
Db	2206	CTGCACTGGGACAGCTCAGCCACAGAGATGGAGTAAAGCAATCCGAAAGGATTT	2265
QY	2316	CTGCTCACTGGCAGAGAGACTGCGATGTGACTGGGAAACCTGCACTGGTGGACACTTACA	2375
Db	2266	CTGCTCCACAGCCACAGGGTGTGTATGGGAGAGAGAG-----ACCAAGT	2310
QY	2376	ACGTTGCTCAATTTATTTCTTTGATATATGCAATTTATATATAGCAATCTGATATAGATGG	2435
Db	2311	ACGTTGCTCAATTTATTTCTTTGATATATGCAATTTATATATAGCAATCTGATATAGATGG	2362
QY	2436	AGCAAACTAGAAATGATGGTGTGATGATGAGACTGGGAAATGTTGAACTGCGAGGGA	2495
Db	2363	AACCAATTTATATGAATTTGATAGCTTAGGCTATGCAAGAAATTTCTGGAATCTCTGAGAG	2422
QY	2496	ACAATTAACAGTAGCTACAGATTTTGGTTTATCTATCTCCAGAGGCCCCACACTCTGCTAGTGA	2555
Db	2423	ATATGAGTTTATACCAAGTGTTTAACTTCTCTCTTACATTTCTCACAC-----TGTTAA	2478
QY	2556	GCCACATCAATACAGAAAGTGAACCTTAAGATGTACACGAAGATG--CCATCCCTCTTTT	2614
Db	2479	GCCATCTCCCATTAAGAGGGCGACTTAAACAAACTGCAAAATGTTTTTGGTTATATCT	2538

[illegible]

FEATURES	COMMENT	JOURNAL	TITLE
source	1. 2417 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM,DB:D030040P06" /db_xref="taxon:10090" /clone="D030040P06" /tissue_type="whole body" /clone_lib="RIKEN Full-length enriched mouse cDNA library" /dev_stage="9 days embryo"	Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/), Tel: 81-45-503-9216)	Group Phase I & II Team. Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2417)
misc.feature	1. 2417 /note="ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STROPLIN-2) homolog [Mus musculus] (SWISSPROT I09DBM0, evidence: FASRY, 92%ID, 96.7%length, match-1796)"	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/	Adachi, J., Alizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukushima, S., Furuno, M., Hanaizumi, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawaji, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohmoto, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sobabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku, Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
BASE COUNT	554 a 695 c 603 g 565 t		
ORIGIN			
Query Match	48.2%; Score 1286.2; DB 11; Length 2417;		
Best Local Similarity	76.1%; Pred. No. 1.3e-308;		
Matches 1696; Conservative	0; Mismatches 483; Indels 50; Gaps 7		
Db	419 CAGGTTGGGAGAGGCTCTTCTAGATGATGATCACTGGCCGAGGTCACGGCCGACAGA 478		
Db	184 CAGCTGGGAGAGGCTCTCTACTACGACGTGATCATCAGGACAGAGCCACGGTGGCAGA 243		
QY	479 TCAAGTTCAGGCGACATGTCATGATGATGAGGACGAGCCAGCTGAGCTGGGAGGAAGT 538		
Db	244 TGAATATCAGCAAAATTTGGATTAATGAGGCAACCCAGTACGCTTGAAGTGGAGAACT 303		
QY	539 GTGTGGCCACAGTGGCCAGGACCAACAGCTGTCTCCCACTGTACTGTGGAGAGACT 598		
Db	304 GCGTTGGCGCATGTGCGGACGATGACCAACATGCTGCGCCCACTGACGCTGACAGAGACCC 363		
QY	599 TGGGCTTCATGAGCCAGATGGGCGTGGCCAGACACTTCTCCAGGCCCCAGAGTACAAAC 658		
Db	364 TGGCTTCATGAGCCAGATGGGCGTGGCCAGACACTTCTCCAGGCCCCAGAGTACAAAC 423		
QY	659 GGGTGGAGAGAGCTGATCGCGAGAGTGGCGGCTTAAAGCACTGGGCTGACACCGCGGAGGCA 718		

Db 424 GGGTGAAGACGTAATGCGGAGCTGGCGGAGTGGCCAAACACGAGATGGCA 483
OY 719 ACATGATGAGGGGGTGTGGGGGGTGAAGGCGAGAGATGACGATTTGGGTGAGC 778
Db 484 ACACGATGATGATGAGGGGTGTGGGGGGTGAAGGCGAGAGATGACGATTTGGGTGAGC 543
OY 779 TCTGTGAAGCCAGGATGCTTATTTCTGACGAGACCCAGCTGCGGGCTGAGAGCTTCA 838
Db 544 TCTGTGAAGCCAGGATGCTTATTTCTGACGAGACCCAGCTTGGGCTTGAAGCTTCA 603
OY 839 CAGCCCAACACTGTGTGAAGACCTTGTCAAGGCTGGCCAAAGGCAACGGCTGTGTCA 898
Db 604 CAGCCCAACACTGTGTGAAGACCTTGTCCGCTGGCCAAAGGCAACGGCTGTGTCA 663
OY 899 TCTGCTGACGAGCCCTGCTGTGACATCTTCAAGGCTGTTGATGTGTCTCTGTATGA 958
Db 664 TCTGCTGACGAGCCCTGCTGTGACATCTTCAAGGCTGTTGATGTGTCTCTGTATGA 723
OY 959 CGTCTGACGAGCCCTGCTGTGACATCTTCAAGGCTGTTGATGTGTCTCTGTATGA 1018
Db 724 CAGTGTGACGAGCCCTGCTGTGACATCTTCAAGGCTGTTGATGTGTCTCTGTATGA 783
OY 1019 TGGGTACCCCTGTGTGTGTGACGATCTGCTGACGATCTTATGTGTGACGATCTGACGA 1078
Db 784 TGGGTACCCCTGTGTGTGTGACGATCTGCTGACGATCTTATGTGTGACGATCTGACGA 843
OY 1079 TTGACGAGGCGGAGAGAGAGAGAGATTTGGCCACGAGAGAGAGAGAGAGAGAGAG 1138
Db 844 TGGACGAG 903
OY 1139 CCCTGTTTGTAGAAAAGTGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 1198
Db 904 CCCTGTTTGTAGAAAAGTGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 963
OY 1199 ATCTGTGAG 1258
Db 964 ATCTGTGAG 1020
OY 1259 CGAGTGTGAG 1318
Db 1021 GAGTGTGAG 1080
OY 1319 TTTCGACGATCTGAG 1378
Db 1081 TTTCGACGATCTGAG 1140
OY 1379 TGTCAATGACATGCGCTTCTCTATTTTGGCCATGAGAGAGAGAGAGAGAGAGAGAG 1438
Db 1141 TGTCCCTCAATGAG 1200
OY 1439 ATACGAG 1498
Db 1201 ATACGAG 1260
OY 1499 TCAATGACGATCTGAG 1558
Db 1261 TGTCTGCAATGATGATGAG 1320
OY 1559 ACACGAG 1618
Db 1321 ACACGAG 1380
OY 1619 ACATGATGATGAG 1678
Db 1381 ACATGATGATGAG 1440
OY 1679 CTTTCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1738
Db 1441 TCTTCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
OY 1739 TGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1798
Db 1501 TGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

OY 1799 ACAATCTCTTCTACCTGCGCGGGGCTTATGATGATGATGATGATGATGATGATGATGATG 1858
Db 1561 ACAATCTCTTCTACCTGCGCGGGGCTTATGATGATGATGATGATGATGATGATGATGATG 1620
OY 1859 CCGCGGAGATTTCCAAAGTGTCTCTCTGCGGGGCTGCTTGTGAAGGCTGATGATGATG 1918
Db 1621 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
OY 1919 AGTTGACGAG 1978
Db 1681 AATTGATGAG 1740
OY 1979 ATAAATCTGATGCTGAG 2038
Db 1741 ACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
OY 2039 TCAATGCGCTGACGCGGCTGCTGATGCTGCTGATGATGATGATGATGATGATGATGATG 2098
Db 1801 TCAATGCGCTGACGCGGCTGCTGATGCTGCTGATGATGATGATGATGATGATGATGATG 1860
OY 2099 AGAAACCAAGTCAAG 2158
Db 1861 AGAAACCAAGTCAAG 1908
OY 2159 GCAAGCCCTTCAACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2218
Db 1909 GACCTTTTCCCGGGCTGCGCCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1968
OY 2219 GACCTTTTCCCGGGCTGCGCCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2278
Db 1969 GATCTGTGAG-----GAG 2021
OY 2279 AGATGAG 2338
Db 2022 AGATGAG 2081
OY 2339 ATGATGAG 2398
Db 2082 ATGATGAG 2120
OY 2399 GAT 2458
Db 2121 --TATATCTTTAAATTAAGAACCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2178
OY 2459 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2518
Db 2179 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2238
OY 2519 TGGCTTCAATCTTCCAGGGGCGCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2578
Db 2239 TAACTTCTCTTCTACATCTTCCACAGC-----TGTAAAGCAGCTCCCAATTAAGAGGCGCA 2294
OY 2579 CCTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2637
Db 2295 CCTAAGAACCAAGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2354
OY 2638 AGCAACGT 2646
Db 2355 CCGCAAGT 2363

RESULT 3
HSM072006
ID HSM072006 standard; RNA; EST; 691 BP.
XX
AC BX481838;
XX
SV BX481838.1
XX
DT 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX

Db	62	CGGGACTTCTAGCTGACTTGGACAGCATGCACAGACGCCACCAAGAAAGCGGAGCTGGCC	121		
QY	1111	ACGAGGGAGAAAGGCTCACTCACTGCGACCCCTGTTTCTAGAAAAAGTCCGTGACTTAGAT	1170		
Db	1122	ACCGTGGAGAAAGGCAACAGCTCTCTGACACCCCTGTTCTCAGAAAAAGTACAGGCGTTGAT	181		
QY	1171	GACTTTCATATGGAAGCGAGAGCGAAGATCTTACAGAGCACTGTGTGGAAGCAGC	1230		
Db	182	GACTTTCGTGTGGAAGGCTGGAGCAAGAGATCTCAACACAGCACCACACAGTACGCTG	241		
QY	1231	GTGACCCCACTAGACACCACTGCTCCCGAGTCCCTACAGAGATGCTGGGGCGGTGCAG	1290		
Db	242	ACCCCTACACAGAGACACTGACTG---TGGAGCTGCTGTGTAGCTGCCCCGGATATATAG	298		
QY	1231	CAGTTTACGAGCGCTGATCCGTGTCAGATTTTCCAAAGCATTTCCGAGACCTGCCACCTC	1350		
Db	299	CAGTTTTCACCCCTGATCCGTGTCAGATTTTCCAAATGATCTCCGGAGCTGCCACCTG	358		
QY	1351	CTCATTCATGAGGGGGAGGCGCTGTCATATGTCATATGACATCCGCTTCCTATTTTGGC	1410		
Db	359	CTCATTCATGAGGGGTGGAGAGCTGCTGATGTCCTCATATGCTTCTCTTACTAGGCG	418		
QY	1411	CATGGGAGCATTCACAGCTCTCCCTTCATGATGATACAGCCGCCCTCTGTTTCATGATGCTGT	1470		
Db	419	CATGGGGGCAAGCAGCTCTCTCTTCATGATGATGACACAGCACCCTCTTCATGATATGAGGCG	478		
QY	1471	CTCATTCCTTTCAACAGCTCATTTCTGATGTCATCTCCAAATGTTACTCAGAGAGGCAATG	1530		
Db	479	CTCATTCCTTTCAATGTCATCTCGAATGTGCTGCCAAATGTCTACTCGAGAGGTCAATG	538		
QY	1531	CTTATAGTATGAAACATGGAAGACGGGCTGTAAACCACTGGTCCATATTTCTTGGCAGATC	1590		
Db	539	CTGTACTATGAGCTGTGGAAGAGGGGCTGT--ACTGCTGTCTTATTTCTTGGCAGATC	596		
QY	1591	CTCGGGAGAGCTTCCGAGAGCAC-TGTGCTATCATATCATATGATGAGGATGCCACCTACTG	1649		
Db	597	CTAGAGAAATGCGCGAGCACTGTGTGCTACGTATCATATCTACGCGATGCCCATCTACTG	656		
QY	1650	GCTGGCCCACTGAGGCGCAGCCCTCAGACCCCTTCCTGCTGC--ACTTCTGCTGTGTGG	1707		
Db	657	GCTGACAAACCTGTGGGCGCCGTGCTGAGCTCTTCTTCTACCACTTCTCTGTGCTG	716		
QY	1708	CTGGTGCTCTTCTGTTGTCAGAGATTAATGCCCCCTGGCGCGGCGCTGCTCCCACTTC	1767		
Db	717	TAGAGAGTCTCTGTCTGTCAGAGACATGAGCCCTTGTGTGCTCTGCCATGTGCTG---CCAACTTC	772		
QY	1768	CACATGCGCTCTCTCTTACAGCAATGCCCTCTACAACTCTTACCT	1814		
Db	773	CACATGCTCTCTCTCTTCTGCA--TGCCTTTAGAAATCTTCTACT	817		
RESULT 5	BP660076	549 bp	RNA	linear	EST 20-DEC-2000
LOCUS	maaz7c08.y1	NCI CGAP L110	Mus musculus	CDNA clone	IMAGE:3812342.5
DEFINITION	similar to TR:Q9VDM4 Q9VDM4 G9964	PROTEIN. ;	RNA sequence.		
ACCESSION	BF660076				
VERSION	BF660076.1	GI:11925210			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus (house mouse)				
REFERENCE	Authors				
FILE	NCI-CPAG	http://www.ncbi.nlm.nih.gov/ncicp.			
JOURNAL	Journal				
COMMENT	Other ESTs: maaz7c08.x1				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabs-remail.nih.gov				
	Tissue Procurement: Jeffrey E. Green, M.D.				
	CDNA Library Preparation: Life Technologies, Inc.				

[illegible]

[illegible]

QY 1471 CTCATCCCTTCAAGCTGATTTGATGTCATCTCCAAATGTTACTCAGAGGCGCAATG 1530
DB 181 CTCATCCCTTCAAGCTGATTTGATGTCATCTCCAAATGTTACTCAGAGGCGCAATG 240
QY 1531 CTTTACTAGAACTGGAAGAGGGGCTGTAGACCACTGGCCATATTTCTTGGCCAAAGATC 1590
DB 241 CTTTACTAGAACTGGAAGAGGGGCTGTAGACCACTGGCCATATTTCTTGGCCAAAGATC 300
QY 1591 CTCGGGGAGCTCCGAGCACTGTGCTTACATCA 1624
DB 301 CTCGGGGAGCTCCGAGCACTGTGCTTACATCA 334

RESULT 9
BB610072 510 bp mRNA linear EST 26-OCT-2001
LOCUS BB610072 RIKEN full-length enriched, adult male liver Mus musculus
DEFINITION CDNA clone 1300007N20 5', mRNA sequence.
ACCESSION BB610072
VERSION BB610072.1 GI:16451685
KEYWORDS EST
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 510)
AUTHORS Arikawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda
'M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okada
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arikawa, T., et al. 2001)
TITLE Unpublished
JOURNAL Contact: Yoshinobu Hayashizaki
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
'M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagii, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
'S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
'Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa
'K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
Location/Qualifiers
1..510
/organism="Mus musculus"
/mol_type="mRNA"

Query Match 12.4% Score 331.2; DB 10; Length 510;
Best Local Similarity 83.7% Pred. No. 2.7e-71;
Matches 375; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1300007N20"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
/clone_11b="RIKEN full-length enriched, adult male liver"
BASE COUNT 142 a 127 c 148 g 93 t
ORIGIN

QY 99 CATGCCGCGAAGGCGGAGAGAGAGGCGCTGCCAAAGGCGCACTCCCGAGGATAC 158
DB 63 CATGCCGCGAAGGCGGAGAGAGAGGCGCTGCCAAAGGCGCACTCCCGAGGATAC 122
QY 159 CTCGGGCTCCAGAGTATGTTCTCTCTGAAAGTACAAAGCCGTGACTTACCTA 218
DB 123 TTCGGGCTCCAGAGAGAGGCGCTGCCAAAGGCGCACTCCCGAGGATAC 182
QY 219 CAGTGGCCAGCCCAACACCTGAGAGTCAAGACCTCACTACAGGTGAGCTGGCCTC 278
DB 183 CAGTGGCTCAAGTCAACACTCTGAGAGTCAAGATCTCACTACAGGTGAGCTGGCCTC 242
QY 279 TCAGTCCCTGTTGAGAGAGGCGCTCACTTCAAGTGCCTGAGATCTCCAGCTG 338
DB 243 TCAGTCCCTGTTGAGAGAGGCGCTCACTTCAAGTGCCTGAGATCTCCAGCTG 302
QY 339 CCAAAATTCCTGTAGAGTGGGATCCAGAACTTAAGCTTCAAGTGAAGTGGGAGAT 398
DB 303 CCAAGACTCCCTGTAGAGTGGGATCCAGAACTTAAGCTTCAAGTGAAGTGGGAGAT 362
QY 399 GCTGCGCATCATAGGAGCTGAGTGTGGAGAGACCTCTTGTAGATGATCACTGG 458
DB 363 GCTGCGCATCATAGGAGCTGAGTGTGGAGAGACCTCTTGTAGATGATCACTGG 422
QY 459 CCGAGGTCAAGGCGGCAAGATCAAGTCAAGGCAATGATGATGAGGAGGAGGAGG 518
DB 423 CAGAGGCGGAGGCGGCAAGATCAAGTCAAGGCAATGATGATGAGGAGGAGGAGG 482
QY 519 GCTGCGCATCATAGGAGCTGAGTGTGGAGAGACCTCTTGTAGATGATCACTGG 546
DB 483 GCTGCGCATCATAGGAGCTGAGTGTGGAGAGACCTCTTGTAGATGATCACTGG 510

RESULT 10
A1157365 511 bp mRNA linear EST 30-SEP-1998
LOCUS A1157365 SUGANO mouse embryo mewa Mus musculus cDNA clone
DEFINITION IMAGE:1885393 5', mRNA sequence.
ACCESSION A1157365
VERSION A1157365.1 GI:3685834
KEYWORDS EST
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Marra, M., Hallier, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Weising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

CHANDLER, J SPENCE, J STOLT, S TAYLOR, G LANG, J SCHEIN, S JONES and

KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 398)
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished

TITLE Contact: Marra M/WashU-NCI Mouse EST Project 1999
JOURNAL Washington University School of Medicine
COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:535908

This read is a RESEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 374.
 Location/Qualifiers

FEATURES
 source

1..398
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:930988"
 /tissue_type="diaphragm"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene mouse diaphragm (#937303)"
 /note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1; EcorI; Site: 2; XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dT. Average insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
 BASE COUNT 89 a 120 c 92 g 97 t
 ORIGIN

Query Match 8.8%; Score 234.6; DB 9; Length 398;
 Best Local Similarity 75.8%; Pred. No. 2.6e-47;
 Matches 304; Conservative 0; Mismatches 94; Indels 3; Gaps 1;

1117 GAGAGGCTCAGTACTGCGACGCCCTGTTCTAGAAAAAGTGGCGTAGTATGACTTT 1176
 1 GAGAGGCGACAGTCTCTTGACAGCCCTGTTCTAGAAAAAGTACAAAGGCTTGATGACTTT 60
 1177 CTATGGAAGCAGAGCAGAGAGTCTTGACGAGGACACCTGTGTGAAAGCAGCGTGACC 1236
 61 CTGTGAAAGCTGAGGCAAGGAATCAACACAAAGCACCACACAGTCAGCCTGACCTC 120
 1237 CCAGTACAGACCACTGCTCCCGAGTCTACGAAGATGCTGGGGCGGTGACAGACTTT 1236
 121 ACACAGGACACTGACTG--TGGGACTGCTGTGAGTGTCCCGGATGATAGACAGTTT 177
 1297 ACGAGCTGATCCGTCGTCAAGATTCCAAAGCACTTCGAGACGTGCCACCTCCTCATC 1356
 178 TCCACGCTGATCCGTCGCAATTTCCAAATGACTTCGCGGACGTGCCACCTGCTCATTT 237
 1357 CATGGGGGGAAGCGCTGTCTATGTCAATGACCAATCGGCTTCTCTATTTTGGCCATGGG 1416
 238 CATGGGTGGAAGCGCTGCTGATGTCCCTCATTCATTTGGCTTCTTACTAGCGCCATGGG 297
 1417 AGCATCCAGCTCTCTTCATGATAGACGCCCTTGTTCATGATGCGTCTCATC 1476
 298 GCCAAGAGCTCTCTTCATGATAGACAGCAGCCCTCTCTTCATGATAGGGGCGCTCATTT 357

QY 1477 CCTTCAAGTCATTCGAGTGTCAATTCACCAATGTACTC 1517
 DB 358 CCTTCAATGTATCTCTGGATGTGCTTCACCAATGTACTC 398

Search completed: July 27, 2003, 08:29:31
 Job time: 4970.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2003, 00:38:04 ; Search time 569.074 Seconds

(without alignments)
9675:628 Million cell updates/sec

Title: US-09-989-981a-7

Perfect score: 2669

Sequence: 1 ggtccctgcctcgcgaac.....catttaaatattatgagc 2669

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2669	100.0	2669	11 US-09-989-981a-7	Sequence 7, Appl1
2	1330	53.6	2019	11 US-09-989-981a-3	Sequence 3, Appl1
3	203.6	7.6	2340	9 US-09-837-992-4	Sequence 4, Appl1
4	203.6	7.6	2340	11 US-09-989-981a-5	Sequence 5, Appl1
5	193.4	7.2	1959	11 US-09-989-981a-1	Sequence 2, Appl1
6	193.4	7.2	2258	9 US-09-837-992-2	Sequence 1, Appl1
7	122	4.6	2025	10 US-09-866-866a-13	Sequence 13, Appl1
8	115.4	4.3	2247	10 US-09-866-866a-26	Sequence 26, Appl1
9	115.4	4.3	2574	11 US-09-961-086-2	Sequence 34, Appl1
10	115.4	4.3	2718	14 US-10-120-687-60	Sequence 60, Appl1
11	115.4	4.3	2718	14 US-10-120-687-60	Sequence 60, Appl1
12	115.4	4.3	2718	14 US-10-120-687-60	Sequence 60, Appl1
13	106.8	4.0	3201	13 US-10-072-621-5	Sequence 9, Appl1
14	105.2	3.9	2930	10 US-09-954-531-591	Sequence 591, App
15	105.2	3.9	2930	14 US-10-171-581-276	Sequence 276, App
16	101.4	3.8	447	10 US-09-960-352-11649	Sequence 11649, A

c	17	93.6	3.5	427	10	US-09-960-352-12839	Sequence 12839, A
	18	93.6	3.5	2400	13	US-10-108-605-244	Sequence 244, App
	19	92.2	3.5	2788	9	US-09-745-763-196	Sequence 196, App
	20	92	3.4	10330	13	US-10-001-189-68	Sequence 68, Appl1
c	21	90.6	3.4	6043	11	US-09-989-981a-9	Sequence 4326, App
c	22	90.2	3.4	377	10	US-09-960-352-4326	Sequence 2262, App
c	23	89.4	3.3	414	10	US-09-960-352-5514	Sequence 5514, App
c	24	88.2	3.3	2223	10	US-09-938-842a-2262	Sequence 6470, App
	25	87.6	3.3	387	10	US-09-960-352-6470	Sequence 3, Appl1
	26	87.2	3.3	1941	14	US-10-090-455-3	Sequence 12, Appl1
	27	87.2	3.3	2687	13	US-10-154-452-3	Sequence 4, Appl1
	28	87.2	3.3	2687	14	US-10-090-455-12	Sequence 7, Appl1
	29	87.2	3.3	3455	13	US-10-072-621-4	Sequence 1148, App
	30	87.2	3.3	3455	14	US-10-090-455-1	Sequence 1, Appl1
	31	86.8	3.3	2687	13	US-10-154-452-7	Sequence 662, App
	32	83.2	3.1	2340	14	US-10-156-761-1148	Sequence 2722, App
c	33	83.2	3.1	9025608	14	US-10-156-761-1	Sequence 4811, App
	34	81.4	3.0	1635	14	US-10-083-357-662	Sequence 6449, App
	35	76.2	2.9	986	10	US-09-764-877-2722	Sequence 4811, App
	36	61.2	2.3	210	9	US-09-294-0938-4811	Sequence 6449, App
	37	57.4	2.2	2529	14	US-10-156-761-1	Sequence 1103, App
	38	57.4	2.2	9025608	14	US-09-938-842a-1103	Sequence 3499, App
	39	56.8	2.1	2118	10	US-10-156-761-3499	Sequence 918, App
	40	56.4	2.1	1221	14	US-10-156-761-4059	Sequence 4059, App
	41	55.8	2.1	3376	14	US-10-037-270-918	Sequence 5662, App
	42	55.2	2.1	774	14	US-10-156-761-5662	Sequence 9027, App
	43	54	2.0	1926	14	US-09-960-352-9027	Sequence 30637, A
	44	51.6	1.9	354	10	US-09-918-995-30637	
	45	51.6	1.9	472	11	US-09-918-995-30637	

ALIGNMENTS

RESULT 1
US-09-989-981a-7
Sequence 7, Application US/0998981a
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Rui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989, 981a
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252, 235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253, 645
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 2669
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (100). (2121)
OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981a-7

Query Match 100.0%; Score 2669; DB 11; Length 2669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GGTGCTCCGCTCCGAGGAACAGAGAGACAGCCCTGGCAGGACAGCTGGCT 60
DB 1 GGTGCTCCGCTCCGAGGAACAGAGAGACAGCTGGCAGGACAGCTGGCT 60

OY	61	AAGAGAGTCGACAGCCGAGGGTCACAGACCTGTGGGCCCATGGCGGGAAGGCGGACAG	120
Db	61	AAGAGAGTCGACAGCCGAGGGTCACAGACCTGTGGGCCCATGGCGGGAAGGCGGACAG	120
OY	121	GAGNAGAGGCTGCCAAGAGGGGCCACTCCCCAGGATACCTCGGGCTCCAGATATGATTG	180
Db	121	GAGNAGAGGCTGCCAAGAGGGGCCACTCCCCAGGATACCTCGGGCTCCAGATATGATTG	180
OY	181	TTCTCTCTGAAAGGACAAACAGCCGTGATCTTACCTACAGTGGGGCAGGCCCAACCCCTG	240
Db	181	TTCTCTCTGAAAGGACAAACAGCCGTGATCTTACCTACAGTGGGGCAGGCCCAACCCCTG	240
OY	241	GAGGTGACAGACCTTCACCTACAGGTGGACCTGGGCTCTCAGAGTCCCTTGTGTATAGAG	300
Db	241	GAGGTGACAGACCTTCACCTACAGGTGGACCTGGGCTCTCAGAGTCCCTTGTGTATAGAG	300
OY	301	CTGGGTCAAGTTCAAAATGCGCTGGACATCTCCAGCTGCAGAAATTTGTGAGCTGGGC	360
Db	301	CTGGGTCAAGTTCAAAATGCGCTGGACATCTCCAGCTGCAGAAATTTGTGAGCTGGGC	360
OY	361	ATCCAGAACCTPAAGCTTCAAAAGTGGAGAGTGGGCGAGATGTCGGCCATATAGGGAGCTCA	420
Db	361	ATCCAGAACCTPAAGCTTCAAAAGTGGAGAGTGGGCGAGATGTCGGCCATATAGGGAGCTCA	420
OY	421	GGTGTGGAGAGCCCTCTTGCTAGATGTGATCACTGGCCGAGGTACAGGGCGCAAGATC	480
Db	421	GGTGTGGAGAGCCCTCTTGCTAGATGTGATCACTGGCCGAGGTACAGGGCGCAAGATC	480
OY	481	AAGTCAGGCCAGATCTGGATCAATGGGCAAGCCAGCTGCCTCAGCTGGTAGAGAAGTGT	540
Db	481	AAGTCAGGCCAGATCTGGATCAATGGGCAAGCCAGCTGCCTCAGCTGGTAGAGAAGTGT	540
OY	541	GTGGGCCCAAGGGCGGCACACACACAGCTGCTCCCACTTGAACATGTGCGAGAGACCTTG	600
Db	541	GTGGGCCCAAGGGCGGCACACACACAGCTGCTCCCACTTGAACATGTGCGAGAGACCTTG	600
OY	601	GCTTCATATGGCCACAGATCGGGCTGCCCAAACTTCTCCAGGGCCACAGTACAAAGG	660
Db	601	GCTTCATATGGCCACAGATCGGGCTGCCCAAACTTCTCCAGGGCCACAGTACAAAGG	660
OY	661	GTGGAGAACGTGATCGCGAGCTGCGGCTTATGGAGCTCGTACACCCGCGTGGGCAAC	720
Db	661	GTGGAGAACGTGATCGCGAGCTGCGGCTTATGGAGCTCGTACACCCGCGTGGGCAAC	720
OY	721	ATGTACGAGCGGGGGTGTGGGGGGGTAGGGGAGAGACAGATTTGGGGTGCAGCTC	780
Db	721	ATGTACGAGCGGGGGTGTGGGGGGGTAGGGGAGAGACAGATTTGGGGTGCAGCTC	780
OY	781	CTGTGGAACCCAGGAATCCTTATTTCTGACGAACCCACTTGTGGGCTGAGACGTTTCACA	840
Db	781	CTGTGGAACCCAGGAATCCTTATTTCTGACGAACCCACTTGTGGGCTGAGACGTTTCACA	840
OY	841	GCCCAACAACCTGTGTAGAGACTTGTGCCAGGCTGGCCAAAGCAACCGGCTGGTCTATC	900
Db	841	GCCCAACAACCTGTGTAGAGACTTGTGCCAGGCTGGCCAAAGCAACCGGCTGGTCTATC	900
OY	901	TCCCTCCACAGCCCTCGCTGTGACATCTTCAAGGCTTTGATCTGTGCTCTCTGATGAG	960
Db	901	TCCCTCCACAGCCCTCGCTGTGACATCTTCAAGGCTTTGATCTGTGCTCTCTGATGAG	960
OY	961	TCGTGGACCCCATCTACTTATAGGGGGGGCCAGCAATGATCTGATATTTTACAGCCATC	1020
Db	961	TCGTGGACCCCATCTACTTATAGGGGGGGCCAGCAATGATCTGATATTTTACAGCCATC	1020
OY	1021	GGCTAACCCCTGTCTCGGTACAGCAATCTGCTACTTCTATGTGTGACTACACAGCAAT	1080
Db	1021	GGCTAACCCCTGTCTCGGTACAGCAATCTGCTACTTCTATGTGTGACTACACAGCAAT	1080
OY	1081	GACAGGCCACAGAGAGACAGAAATTTGGCCACAGGGAGAAAGGCTCACTGCGACAC	1140
Db	1081	GACAGGCCACAGAGAGACAGAAATTTGGCCACAGGGAGAAAGGCTCACTGCGACAC	1140
OY	1141	CTGTCTTACAAAAAGTGCCTGACTTATGATGATCTTCTATGTGAAAGCGAGAACGAAAGAT	1200

Db	1141	CTGTTCTAGAAAGAGTGGTGACTTAGTATGACTTTCTATGGAAGAGACAGACGAAAGAT	1200
Qy	1201	CTTGACAGAGACACTGTGTGGAAGACAGCGTGACCCACTAGACACCACTGCTCCCG	1266
Db	1201	CTTGACAGAGACACTGTGTGGAAGAGACAGCGTGACCCACTAGACACCACTGCTCCCG	1266
Qy	1261	ACTCTAGGAACATGCTGTGGGGGGGTGAGAGATTAGAGCGCTGATCCGTGCTAGATT	1322
Db	1261	ACTCTAGGAACATGCTGTGGGGGGGTGAGAGATTAGAGCGCTGATCCGTGCTAGATT	1322
Qy	1321	TCCACAGACTTCCGAGACTGCGCCACCCTCCATCCATGAGGGGCGAGAGGCTGTGTGAT	1380
Db	1321	TCCACAGACTTCCGAGACTGCGCCACCCTCCATCCATGAGGGGCGAGAGGCTGTGTGAT	1380
Qy	1381	TCAATGACCACTGCGCTTCCTCTATTTTGGCCATGGGAGACATCCAGCTCTCTTCATGGAT	1440
Db	1381	TCAATGACCACTGCGCTTCCTCTATTTTGGCCATGGGAGACATCCAGCTCTCTTCATGGAT	1440
Qy	1441	ACAGCGGCGCTCTTCTTCATGATGCGGTGCTCTCATCCCTTCACAGCTATTCTGGATGTC	1500
Db	1441	ACAGCGGCGCTCTTCTTCATGATGCGGTGCTCTCATCCCTTCACAGCTATTCTGGATGTC	1500
Qy	1501	ATCTCCAATGTTTACTCAGAGAGGGGCATAGCTTTACTATGAACTGGAAGAGGGGCTATAC	1560
Db	1501	ATCTCCAATGTTTACTCAGAGAGGGGCATGCTTTACTATGAACTGGAAGAGGGGCTGTATC	1560
Qy	1561	ACCAGTGGTCCATATTCTTCTGCCAAAGATCCTCGGGAGACTTCGGAGACACTGTGCTATC	1620
Db	1561	ACCAGTGGTCCATATTCTTCTGCCAAAGATCCTCGGGAGACTTCGGAGACACTGTGCTATC	1620
Qy	1621	ATCATCATCTTACGGGATGCGCCACCTACTGTGCTGGCCAACTGAGGCCAGGCTCTCAGCCC	1680
Db	1621	ATCATCATCTTACGGGATGCGCCACCTACTGTGCTGGCCAACTGAGGCCAGGCTCTCAGCCC	1680
Qy	1681	TTCTGCTGCACTTCTCTGTGTGTGGTGGTGGATCTTCTGTGAGAGATTATGGCCCTGTG	1740
Db	1681	TTCTGCTGCACTTCTCTGTGTGTGGTGGTGGATCTTCTGTGAGAGATTATGGCCCTGTG	1740
Qy	1741	GCGCGCGGCGGCGCTGCTCCGCCACCTTCCACATGGGCTCTTCTTCAGCAATGCCCCCTATAC	1800
Db	1741	GCGCGCGGCGGCGCTGCTCCGCCACCTTCCACATGGGCTCTTCTTCAGCAATGCCCCCTATAC	1800
Qy	1801	AACCTCTTCTACCTTCGCGGGGGGCTCTATGATTAACCTTGAGACAGCTCTGGACAGTGGCC	1866
Db	1801	AACCTCTTCTACCTTCGCGGGGGGCTCTATGATTAACCTTGAGACAGCTCTGGACAGTGGCC	1866
Qy	1861	GCGTGGATTTCCAAAGTGTCTCTGCTCGGGGGGATGTTTGAAGGGCTGATGAAGTTTACG	1920
Db	1861	GCGTGGATTTCCAAAGTGTCTCTGCTCGGGGGGATGTTTGAAGGGCTGATGAAGTTTACG	1920
Qy	1921	TTTCAGCAGAGAACTTATTAATATGCTCTCGGGGAACTCAGCATTCGCGGTCTCAGAGAT	1980
Db	1921	TTTCAGCAGAGAACTTATTAATATGCTCTCGGGGAACTCAGCATTCGCGGTCTCAGAGAT	1980
Qy	1981	AAATATCCATGCGCATGGAGCTGAGACACGTACACCTCTCTACGGCATCTACCTATGCTC	2040
Db	1981	AAATATCCATGCGCATGGAGCTGAGACACGTACACCTCTCTACGGCATCTACCTATGCTC	2040
Qy	2041	ATTGGCTCTCAGCGGTGCTTACATGGTCTCTGACTACAGCTGCTTAAAGTTTATCAAAACAG	2100
Db	2041	ATTGGCTCTCAGCGGTGCTTACATGGTCTCTGACTACAGCTGCTTAAAGTTTATCAAAACAG	2100
Qy	2101	AAACCAAGTCAAGACTGATGATTCACGCGCAGACACTGTGCGCGGTGTGGGGAGACTGAGC	2160
Db	2101	AAACCAAGTCAAGACTGATGATTCACGCGCAGACACTGTGCGCGGTGTGGGGAGACTGAGC	2160
Qy	2161	AGACCTTCAACTGACATCCCTCTCTCAGAGAGCCCTCTTCGTGGGAGACATGAGCAATGA	2220
Db	2161	AGACCTTCAACTGACATCCCTCTCTCAGAGAGCCCTCTTCGTGGGAGACATGAGCAATGA	2220
Qy	2221	CCCTACAGATCTGAGTACATCCGCGCCAGAGGTGTGACAGTGGCAGACACCACTCAG	2280
Db	2221	CCCTACAGATCTGAGTACATCCGCGCCAGAGGTGTGACAGTGGCAGACACCACTCAG	2280

Db 2221 CCTACAGATGCTACAGCTACATCCGAGGAGGCTGCTGAGTGCACAGACACGACCCAG 2280
 QY 2281 GATGCGAGTAGAATAAGACAGTCAAGAGGATTTCTGCTACCTGGCAGAGACCTCGAT 2340
 Db 2281 GATGCGAGTAGAATAAGACAGTCAAGAGGATTTCTGCTACCTGGCAGAGACCTCGAT 2340
 QY 2341 GACTGGAGAGAAACCTGCACCTGCTGGGACACCTAACAGTGTATTTATTTCTTTTGA 2400
 Db 2341 GACTGGAGAGAAACCTGCACCTGCTGGGACACCTAACAGTGTATTTATTTCTTTTGA 2400
 QY 2401 TATGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2460
 Db 2401 TATGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2460
 QY 2461 CTAGACTGTGAGAGATTTGTTGAACTGGAGGACATTAACAGTGTATTTATTTATTTAT 2520
 Db 2461 CTAGACTGTGAGAGATTTGTTGAACTGGAGGACATTAACAGTGTATTTATTTATTTAT 2520
 QY 2521 GCTTCATCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2580
 Db 2521 GCTTCATCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2580
 QY 2581 TTAGAGTGTACAGAGAGATGCTATCTTTCTTTTGTGTGGGTCTATGGGCTCCAAAGC 2640
 Db 2581 TTAGAGTGTACAGAGAGATGCTATCTTTCTTTTGTGTGGGTCTATGGGCTCCAAAGC 2640
 QY 2641 CAACGTGACAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2669
 Db 2641 CAACGTGACAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2669

RESULT 2

US-09-989-981a-3
 : Sequence 3, Application US/09989981A
 : Publication No. US20030049730A1
 : GENERAL INFORMATION:
 : APPLICANT: Hobbs, Helen H.
 : APPLICANT: Shan, Bel
 : APPLICANT: Barnes, Robert
 : APPLICANT: Tian, Hui
 : APPLICANT: Tularek Inc.
 : APPLICANT: Board of Regents, The University of Texas System
 : TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 : CURRENT APPLICATION NUMBER: US/09/989,981A
 : CURRENT FILING DATE: 2002-07-23
 : PRIOR APPLICATION NUMBER: US 60/252,235
 : PRIOR FILING DATE: 2000-11-20
 : PRIOR APPLICATION NUMBER: US 60/253,645
 : PRIOR FILING DATE: 2000-11-28
 : NUMBER OF SEQ ID NOS: 13
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 3
 : LENGTH: 2019
 : TYPE: DNA
 : ORGANISM: Mus musculus
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (1)..(2019)
 : OTHER INFORMATION: mouse ABCG8 (mABCG8)
 : US-09-989-981a-3

Query Match 53.6%; Score 1430; DB 11; Length 2019;
 Best Local Similarity 82.0%; Pred. No. 0;

Matches 1059; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

QY 100 ATGGCCGGGAGAGCGGAG 159
 Db 1 ATGGCTGAG 60
 QY 160 TCGGCGCTCCAGAGATAGATTTCTCTCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
 Db 61 TCGGCGCTCCAG 120

QY 220 AGTGGCAG 279
 Db 121 AGTGGCAG 180
 QY 280 CAGGTCCTTGGTTGAG 339
 Db 181 CAGGTCCTTGGTTGAG 240
 QY 340 CAGATTTCTTGTAG 399
 Db 241 CAGATTTCTTGTAG 300
 QY 400 CTGGCCATATAG 459
 Db 301 CTGGCCATATAG 360
 QY 460 CGAGGTGACAGGAG 519
 Db 361 AGAGGCGCAGGAG 420
 QY 520 CCTCAGCTGTGAG 579
 Db 421 CCTCAGCTGTGAG 480
 QY 580 TTGACTGTGAG 639
 Db 481 TTGACTGTGAG 540
 QY 640 CAGGCCAGGAG 699
 Db 541 CAGGCCAGGAG 600
 QY 700 GCTGACACCCGCGTGGGAG 759
 Db 601 GCTGACACCCGCGTGGGAG 660
 QY 760 GTACAGATTTGGGAG 819
 Db 661 GTACAGATTTGGGAG 720
 QY 820 TCTGGGCTGAG 879
 Db 721 TCTGGGCTGAG 780
 QY 880 GGCAGACCGGCTGCTCATCTCTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
 Db 781 GGCAGACCGGCTGCTCATCTCTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 QY 940 GATCTGTCTCTCTGATGAG 999
 Db 841 GATCTGTCTCTCTGATGAG 900
 QY 1000 GTCCAGATTTACAG 1059
 Db 901 GTCCAGATTTACAG 960
 QY 1060 TATGTGAG 1119
 Db 961 TATGTGAG 1020
 QY 1120 AAGGCTGAG 1179
 Db 1021 AAGGCTGAG 1080
 QY 1180 TGGAG 1239
 Db 1081 TGGAG 1140
 QY 1240 CTAGACACCACTGCGCTCCGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299
 Db 1141 CAGGACACTGACTG---TGGAGACTGCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197

QY	1300	ACGCGTAACTCCGTCGACATTTCCAAAGACATCCGAGACACCGCCACCCTCCATCCAT	1353
Db	1198	ACCGTGAATCCGTCGACATTTCCAAATACTCCGGGACACCGCCACCCTCCATTCAT	1257
QY	1360	GGGGCGAGACCCCTGTCTGATGTCAATGACATCGGCTTCCTFATTTTGGGCATGGGAGC	1419
Db	1258	GGGTGCGAAGCCCTGCTGATGATTCCTGCTCATATTTGGCTTCCTTTACTACGGCCATGGGCC	1317
QY	1420	ATCCAGCTCTCTTCATGAGATACAGCGCCCTCTTGTTCATGATCGGTGCTCTACCCCT	1479
Db	1318	AAGAGGCTCTCTTCATGAGACACGAGCCCTCCCTTCATGATAGGGGGCCCTCATTCCT	1377
QY	1480	TTTCAAGCATTTGGAATGATCTCCAAAGSTTACACAGAGAGGAGGCAATCTTACTAT	1539
Db	1378	TTCAATGTCATCTGAGATGCTGTCCAAATGTCACTCGAGAGGTCAATCTGTACTAT	1437
QY	1540	GAACGTGAGAGACGGGCTGTACACCACTGTCATATTTCTTGGCAAGATCCTCGGGGAG	1599
Db	1438	GAGCTGGAAGACGGGCTGTACACGTGCTGATTCCTTATTTCTTGGCAAGATCCTGAGAGA	1497
QY	1600	CTTCCGGAGACACTGTGCTTACATATCATCTTACGGGATGCCACCTACTGGCTGGCAAC	1659
Db	1498	TTGGCGGAGCACTGTGCTTACGTCATCATCTACACGATGCCACATCTACTGGCTCAACAAC	1557
QY	1660	CTGAGAGCAGAGCCCTCCAGACCCTTCTGCTGACATTCGCGGGGTGGGCTGGTCTTC	1719
Db	1558	CTGGGGCCCGGCTGAGCTTCTTCTTCTTACACTTCTGCTGCTGTGGTGGTGGTCTTC	1617
QY	1720	TGTTTCAGAGATTAATGCGCTTGGCGCGCGCGGCGCTGTCCCACTTCACATGAGCCTCC	1779
Db	1618	TGCTGCGAGGACCAATGAGCCCTGGCGCTGCTCTGACCATGCTGCCACCTTCACATGTCCGCC	1677
QY	1780	TTCTTCAGCAATGGCCCTTACAAACTCTCTTCACTCCGCGGGGCTTCATGATTAACCTTG	1839
Db	1678	TTCTTCAGCAATGGCCCTTACAAACTCTCTTCACTTACTGCGGGCTTCATGATTAACCTTG	1737
QY	1840	AGCAGCCTGTGGACAGTGGCCCGCGCTGTGATTTCCAAAGTCCCTTCCGCGGTGGTGT	1899
Db	1738	GACAACTGTGGATGATGTGCTGCATGTGATCTCCAAAGTGTGTTCCCTCCGCTGGTGTCTC	1797
QY	1900	GAAAGGCTGATGAAGATTCAAGTTCAGACAGAAAGATTATATAATGCTCTTCGGGAACCTC	1959
Db	1798	TCGGGCGTGAATGACGATTAATTAATTAATGACACCTTTTACACACACAAATAGCGACAACTTC	1857
QY	1960	ACCATCGCGTCTTCAGAGATTAATAATCTCGATGCCATGAGAGCGTGAACCTCCCTC	2019
Db	1858	ACCTTCTTCATCTCCGAGACACGATATATCAGTCCATGGACCTGAACTGCGACATCCACTC	1917
QY	2020	TACGCATCTACCTCATGTCATTTGGCTCAGCGGCTGCTTCATGATGCTCTGTACTAGTGTG	2079
Db	1918	TATGCGATCTAACCTCATTTGTCATCGGATCAGGTACGAGCTTCCTCGTTCGTACTACTTA	1977
QY	2080	TCCTTAAGTTCATCAACAAGAAACCAAGTCAAGACTGGTGA 2121	
Db	1978	TCCTTGAAGTCAACAACAAGTCAATTCACAGAGGTTGA 2019	

~~RESULT~~ 3

```

: Sequence 4 Application US/09837992
: Patent No. US20020081687A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Tian, Hui
: APPLICANT: Schultz, Joshua
: APPLICANT: Shan, Bei
: APPLICANT: Tularik Inc.
:
: TITLE OF INVENTION: Slicosterolemia Susceptibility Gene (Ssc): Compositions
:
: TITLE OF INVENTION: And Methods of Use
:
: FILE REFERENCE: 018781-006020US
:
: CURRENT APPLICATION NUMBER: US/09/837,992
:
: CURRENT FILING DATE: 2001-04-18
:
: PRIOR APPLICATION NUMBER: US 60/198,465
:
: PRIOR FILING DATE: 2000-04-18

```

```

? PRIOR APPLICATION NUMBER: US 60/204,234
? PRIOR FILING DATE: 2000-05-15
? NUMBER OF SEQ ID NOS: 45
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 4
? LENGTH: 2340
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: human sitosterolemia gene (SSG)
? NAME/KEY: CDS
? LOCATION: (107)..(2062)
? OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
? OTHER INFORMATION: protein
US-09-837-992-4

```

Query Match	7.68;	Score 203.6;	DB 9;	Length 2340;
Best Local Similarity	54.48;	Pred. No. 1.2e-50;		
Matches 432; Conservative	0;	Mismatches 359;	Indels 3;	Gaps 1

QY	335	GCTGCCAGAAATTTCTGTAGCTGGGCGATCCAGAAACCTTAAGCTTCAAACTGAGAACTGGGC	394
Db	285	GCCTGCAGCAATGGACCGACGACAGATCCCTCAAAAGATGTCTCTTTGACTGGAGACGGGC	344
QY	395	AGATGCTGGCCATCATAGGAGCTCTAGGTTGTGGAGACCTCTCTTCTAGATGTGATCA	454
Db	345	AGATCATGTGATCTCTAGGAGCTCCAGCTCCGGGAAAACACGCTGCTGGACGCCATGT	404
QY	455	CTGGCCCAAGCTACGGGGCGCAAGATCAAGTCAAGCCCAATCTGGATCAATGGGACGCCA	514
Db	405	CCGGGAGGCTGGGGCGCGCGGGGACCTTCTCGGGGAGGTGTATCTGAACGGCGGGCGC	464
QY	515	GCTGCGCTCAGCTGGTGAAGATGTGGGCCCAAGTGCAGCAACACGAGCTGTC	574
Db	465	TGGCCCGGAGCAATTCAGACATCTCTCTCTAGTCTCTGACAGGACACACCTGCTGTA	524
QY	575	CCAACTTGACTGTGCGAGAGACCTTGGCTTATTTGCCAGATGGCGGTGCCCAAACT	634
Db	525	GCACCTTCACCGTCCGGGAGACGCTGCACCTACCGGCGCTGTGGCCATCCGGCCGGCA	584
QY	635	TCCTCCAGGCCCAAGCTGTACAAAGGGTGGAGACGTGATGCGGAGCTGGGGCTTAAGC	694
Db	585	ATCCCGGCTCTTCC---AGAGAAGGGGGAGGGCGCTATGACAGAGCTGAGTCTGAGCC	641
QY	695	AGTGCCTGCACCCCGGTGGGCAACATGTACGTGCGGGGGTTGTGGGGGGGTGAAGCCA	754
Db	642	ATGTGGAGAACGCACTGATTTGGCAACTACACTTTGGGGGCAATTTCCACGGGTGACGGC	701
QY	755	GGAGATCAGCAATTTGGGGTGCAGCTCTGTGGAAACCCAGAACTCTTAATTTCTGACGAA	814
Db	702	GCCGGGCTCTCATCGACAGCCAGCGTGCAGAGATCTTAAGGTCAATGCTGTTGATGAGC	761
QY	815	CCACCTGTGGGTCGACAGCTTCAACAGCCCAACCTGGTAAAGACCTTTGTCAGGCTGG	874
Db	762	CAACCAAGCGCTGGAGTCATGACTCTTAATTCAGATTTGCTCTCTCGGTGGAACTGG	821
QY	875	CCAAAGCAGCCGCTGGTCTCTCTCTCTCCCAACAGCGCTCGCTGACATCTTCAAGGC	934
Db	822	CTGGCAGAGAACCGAATTTGTGTTCTACACATTCACAGCCCGCTTCTGAGCTTTTTCAGC	881
QY	935	TGTTTGATCTGTCTCTCTGATGACGTGTGGACCCCACTTACTTAAAGGGCGGCCAGC	994
Db	882	TCATTGCAAAATTTGCCATCTCTGAGCTTGGAGAGTGAATTTCTGTGGACGCCAAGGG	941
QY	995	ACATGGCCGAGTATTCACAGCCATCGGGCTACCCCTTCTCTCGCTACAGCAATCTGCTG	1054
Db	942	AAATGCTGATTTCTTAATGACTGCGGTTAACCTTCTCTGAACTTCAAACTCTTTTG	1001
QY	1055	ACTTCTATGTGAGCTGACAGCAATTACAGGGCGACAGAGAGACAGAAATTTGGCCAGCA	1114
Db	1002	ACTTCTATATGAGACTGACGTCAAGTGAATACCCAAAGCAAGAGACGGGAAATGAAACT	1061
QY	1115	GGGAGAGGCTCAG	1128

Query Match	7.28;	Score 193.4;	DB 9;	Length 2258;
Best Local Similarity	53.48;	Pred. No. 1.5e-47;		
Matches 429;	Conservative 0;	Mismatches 371;	Indels 3;	Gaps 1;

OY	335	GCTGCAGAAATTTCTTGAGAGCTGGGCAATCCGAAACCTAAGCTTCAAAAGTGAAGAGTGGCC	394
Db	228	GCACGCAAAATGTGGAGCAGGCAAAATCCCTCAAAAGTGTCTCTGTGTACATCGAGAGTGGCC	287
OY	395	AGATGCTGGCCATCATAGAGGAGCTCAGAGTTGTGGAGAGGCTCCCTGCTGTAGATGTATCA	454
Db	288	AGATTATGTGATCTTTAGGACGCTCAGAGGCTCAGGGAAGACCAGCTCTGGACGCCATCT	347
OY	455	CTGGCCGAGGCTCACGGGGCAAGATCAAGTCAGGCCAGATCTGGATCAATGGGCAGCCA	514
Db	348	CCGGGAGGCTGCGGGCACTGGGACCCCTGGAAAGGGGAGGTGTGTATATGGCTGCGAGC	407
OY	515	GCTGCGCTCAGCTGTGTAGGAAGTGTGGGCCCCAGCGGCCACACACACAGCTGTCC	574
Db	408	TGCGCAGGAGCAATTCACAGACTGCTTCTCTTACGTCTGCAAGCAGCACTTTTCTGA	467
OY	575	CCAACTTGACTGTGCGAGAGACCTTGGCCCTTCATTTGCCAGATGCGGCTGCCAGAACT	634
Db	468	GCACCTTCACCTGTGCGGAGACCTTGGGATACAGC--GATGCTGGCCCTCTGCGGCA	524
OY	635	TCTCCAGAGCCCAACCGTGCACAAAGGGTGGAGAGACGTGATGCGGAGCTGCGGCTTAA	694
Db	525	GCTCCGCGGAACTTCTACACAGAAAGGATAGAGGAGCATGACAGACTAGGCTTGAGCC	584
OY	695	AGTCCGCTGACACCCGGTGGGCAACATGTACGTGCGGGGGTTGTGCGGGGGTGAAGGCA	754
Db	585	ACGTGGGGGACCAATATATGTGGCAGCTATATATTTTGGGGGAATTTCCAGTGGCGAGCGCC	644
OY	755	GGAGAGTCAGCAITGGGGTGCACTCTGTGGAAACCCAGGAATCTTATTTCTGCAGCAAC	814
Db	645	GCCAGTTTTCATGCGACGCCAACTCCTTAGAGACCCTCAAGTCATGATCTGATGATAGC	704
OY	815	CCACTCTGGGCGTGCAGAGCTTCAACAGCCCAACACCTGTGTGAAGACTTGTCCAGGCTGG	874
Db	705	CAACCAAGAGCTGAGCTGATGATGCTCAAAATCAAAATTTGCTTCTTGGCTGGAGCTGG	764
OY	875	CCAAAGGCAACCGGCTGTGCTCATCTATCTCCCTGCACACAGGCTCGCTGACATCTTCAGGC	934
Db	765	CTGCGAGGAGCCGAATTTGTGATTGTCAACATTCACACAGCTCTGCTCTGAGACTCTTCCAAC	824
OY	935	TGTTTGAATCTGTGCTCTGATGACGCTGTGGCAACCCCAATCTAATTAGGGGCGCCAGC	994
Db	825	ACTTGCACAAATTTGCCATCTCTGATCTTACGAGAGTGTGTGTTCTGTGGCACCCCAAGG	884
OY	995	ACAATGGCCAGTATTTTACAGAGCCATGGGTCACCCGTGCTGCTCAAGCAATCTGCTG	1054
Db	885	AGATGCTTGGCTTCTTCAATTAAGTGTGTATACCCCTGTCTTGAACATTTCCAAATCCCTTGG	944
OY	1055	ACTTCTATGTGAGCTTGACCAAGCATTTGACAGGGCGGACAGAGACAGAGATTTGGCCACA	1114
Db	945	ATTTTTTCATGAGACTTGACATCTAGTGACACCCAAAGACAGAGCGGAAATGAAACGT	1004
OY	1115	GGGGAAGGCTCACTGACTGCA	1137
Db	1005	ACAAAGCAGTACAGATGCTGGAA	1027

```

RESULT 7
US-09-866-866A-13
: Sequence 13, Application US/09866866A
: Patent No. US2002010224A1
: GENERAL INFORMATION:
:
: APPLICANT: Sorrentino, Brian
: APPLICANT: Schuetz, John
: TITLE OR INVENTION: A Method of Identifying and/or Isolating Stem Cells
: FILE REFERENCE: 1340-1-021CIP2
: CURRENT APPLICATION NUMBER: US/09/866,866A
: CURRENT FILING DATE: 2001-08-30

```

; PRIOR APPLICATION NUMBER: 09/584,586
 ; PRIOR FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: PCT/US99/11825
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: 60/086,988
 ; PRIOR FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 2025
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-866-866A-13

Query Match 4.68; Score 122; DB 10; Length 2025;
 Best Local Similarity 51.78; Pred. No. 5.8e-26;
 Matches 278; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 538 TGTGTGCCCCCGTGGCCAGCACAACACCTGCTCCCACTGAGTGTGGAGAGACC 597
 DB 374 TGTTCAGGTATATGTTGTTCAAGATGACGTGTGATGGGACCTGACAGTGAAGAAAC 433
 QY 598 TTGGCCTTCATTGCCAGATGGCGTCCAGAACCTTCGCCAGGCCGCGTGACAAA 657
 DB 434 TTACAGTTCACAGCAGCTCTTCGACCTCCAAACAATATGAAGATCATGAAAAAATGAA 493
 QY 658 AGGGTGGAGAGACTGATCCGGAGCTGGCGCTTAAAGCAGTGGCCTGACCCGCGTGGGC 717
 DB 494 CGGATTACACATCAATCTTAAAGATGAGTGTGAAAAAGTACCAATTTCTAAGGTCGGA 553
 QY 718 AACATGTAGTGGCGGGGTTGTGCGGGGTGAGCGAGAGAGTACATGAGTGGGGGCGAG 777
 DB 554 ACACATTTATCCGTTGTCGATCTGTGAGAGAGAAAGAAAAGCAAGCATAGGATGAGAG 613
 QY 778 CTCTCTGGAAACCCAGGAATCTTATTCGAGCAACCAACCTCTGGGCTCGACAGCTTC 837
 DB 614 CTGATCACTGACCTTCATCCCTCTCTCTGATGATGACCCAGCAGCTGTTTGGACTCAAG 673
 QY 838 ACAGCCCAACACTGGTGAAGACTGTCCAGGCTGGCCAAAGGCAACGGCGTGGCTC 897
 DB 674 ACAGCAATGCTGTCTCTTTGCTCTGAAAAGATGCTTAAACAGGGTGAACATCATC 733
 QY 898 ATCTCCCTCCACAGGCTGCTGTGATCTTCAGGCTGTTGATCTGGTCCCTCGATG 957
 DB 734 TTCTCATTCATCAGGCTGCTGTGATCTTCAGGCTGTTGATCTGGTCCCTCGATG 793
 QY 958 ACCTGTCGACCCCACTTACTTAAAGGGGGGCCAGACATGATGATTTTCAACAGCC 1017
 DB 794 GCTTCGGGAAACTGCTGTTCATGATGGCCAGACAGAAAGCCCTGAGATCTTGCATCA 853
 QY 1018 ATCCGTACCCCTGCTGCTGCTACAGCAATCTCTGCTGACTTCTATGTGACCTGACCA 1075
 DB 854 GCAGGTACCACTGTGAGCCCTTAAACAACCCCTGGGATTTTCTTGATGTCAATCA 911

RESULT 8

; Sequence 26, Application US/09866866A
 ; Patent No. US20020102244A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sorrentino, Brian
 ; APPLICANT: Schuetz, John
 ; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
 ; FILE REFERENCE: 1340-1-021CIP2
 ; CURRENT APPLICATION NUMBER: US/09/866,866A
 ; PRIOR FILING DATE: 2001-08-30
 ; PRIOR APPLICATION NUMBER: 09/584,586
 ; PRIOR FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: PCT/US99/11825
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: 60/086,988
 ; PRIOR FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 26
 ; LENGTH: 2247
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-866-866A-26

Query Match 4.38; Score 115.4; DB 10; Length 2247;
 Best Local Similarity 51.28; Pred. No. 6e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 548 ACCTGCCCCAGCACAACACCTGCTCCCACTGAGTGTGGAGAGACTTGGCCTTCA 607
 DB 561 ACCTGTCACAGATGATGATGTTGTGTTGGCACTGACGCTGAGAGAAACTTACACTTCT 620
 QY 608 TTGCCAGATGCGGCTGCCAGAACCTTCTCCAGGGCCAGCGCTGACAAAAGGTGAGG 667
 DB 621 CAGCAGCTCTTCGGCTTGCACAACATGATGACATGATGATGATGATGATGATGATGAT 680
 QY 668 ACCTGATCGCGAGAGCTGGCTTAAAGCAGTGGCTGACACCCGCTGGGCAACATGACG 727
 DB 681 GGGTCATTCAGAGTTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 740
 QY 728 TGGGGGGGTTGCGGGGGGTGAGCGCAGAGAGTCAACATTTGGGTTGACCTCTCTGGA 787
 DB 741 TCGGTGTGTGTGGAGGAGAAAGAAAGAGACTGATATGATGATGATGATGATGATGATGAT 800
 QY 788 ACCAGGATCTCTTATCTGAGAACCCACCTGCGGCTGACAGCTTACAGCCACCA 847
 DB 801 ATCTCTCATCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 860
 QY 848 ACCTGTGAAGACCTTGTCCAGGCTGGCCAAAGCAACCGCTGGTCTCATCTCCCTCC 907
 DB 861 CTGTCTTTTGTCTCGAAGAAAGATGCTTAAAGCAGGAGAGCAACATCATCTTCTCATTC 920
 QY 908 ACCAGCTGCTGCTGATCTTATGAGCTGTTTATGCTGCTGCTGATGAGCTTGGCA 967
 DB 921 ATCAGCTGATATTCATCTTCAAGCTTGTATATGACCTTATTTGGCCCTCAGGAA 980
 QY 968 CCCCCATCTACTTAAAGGGGGGCCAGACATGCTCCAGTATTCACAGCATCGGCTACG 1027
 DB 981 GACTTATGTTCCAGGGGCTGCTCAGAGAGCCCTTGGATATCTGATATCACTGCTTATC 1040
 QY 1028 CCGTCTCTGCTACAGCAATCTGCTGACTTCTATGTGAGACTGA 1072
 DB 1041 ACTGTAGGCGCTTATATTAACCTGCAAGACTTCTTCTTGGAGATCA 1085

RESULT 9

; Sequence 2, Application US/09961086
 ; Publication No. US2003003645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
 ; APPLICANT: ROSS, Douglas D.
 ; APPLICANT: DOYLE, L. Austin
 ; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
 ; FILE REFERENCE: EP19376-019
 ; CURRENT APPLICATION NUMBER: US/09/961,086
 ; PRIOR FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 60/073,763
 ; PRIOR FILING DATE: 1998-02-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/02577
 ; PRIOR FILING DATE: 1999-02-05
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2418
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-961-086-2

Query Match 4.3%; Score 115.4; DB 11; Length 2418;
 Best Local Similarity 51.2%; Pred. No. 6.3e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

548 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGCTGGCCCTCA 607
 606 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGCTGGCCCTCA 665
 608 TTGCGCAGATGGCGGCTGCCAGAACCTTCTCCAGAGCCAGGCTACAAAAGGTGAGG 667
 666 CAGCAGCTCTTGGGCTTGCACAACTATGACGAATCATGAAAAACGACAGGATTAACA 725
 668 ACCTGATCGCGAGCTGGCGCTTGAAGCAGTGGCTGACACCCGCTGGGCAACATGACG 727
 726 GGGTCATTCAGAGATGAGTCTGATTAAGTGAAGTGGCAGACTCCAAAGGTGGAACTCAGTTTA 785
 728 TCGCGGGGTGTGCGGGGGGTGAGCGCAGAGAGTACATGCGGTGACAGCTCCCTGGA 787
 786 TCCGAGGTGTGTCTGAG 845
 788 ACCAGGAATCTTATTCGACAGCAAGACCCAGCTGCGCTGACAGCTTACAGCCACA 847
 846 ATCTTCATCTTGTCTTGATGAGCCTACAGCTGCGCTTACAGCTTACAGCTTACAG 905
 848 ACCTGATGACAGCTTGTGACAGCTGGCCAAAGGACCGGCTGCTCATCTCCCTCC 907
 906 CTGCTCTTCTCTCTGAAAAGAGATCTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 965
 908 ACCAGCTGCGCTGACATCTTCAAGCTGCTTGAATCTGCTCTCTGATGACGTGGCA 967
 966 ATCAGCTGAGATCTTCAATCTTCAAGTGTGATGAGCTTCACTTATTTGGCCTCAAGAA 1025
 968 CCCCCATCTACTAGGAGGCGGCGCCAGACATGCTCAGATTTTCACAGCCATCGGCTACG 1027
 1026 GACTTATGTTCCAGAGGCGCTCTCAGAGAGGCGCTTGGATTTGATGATCAGCTGTTATC 1085
 1028 CCTGCTCTGCTTACAGCAATCTGCTGATCTTATGAGACCTGA 1072
 1086 ACTGTGAGGCTTATATATACCTGACACTTCTTCTGACATCA 1130

RESULT 10

US-09-981-353-34
 : Sequence 34, Application US/09981353
 : Patent No. US20020160382A1
 : GENERAL INFORMATION:
 : APPLICANT: Lasek, Amy W.
 : APPLICANT: Jones, David A.
 : TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 : FILE REFERENCE: PA-0038 US
 : CURRENT APPLICATION NUMBER: US/09/981.353
 : CURRENT FILING DATE: 2001-10-11
 : NUMBER OF SEQ ID NOS: 194
 : SOFTWARE: PERL Program
 : SEQ ID NO 34
 : LENGTH: 2574
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc.feature
 : OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CBI
 US-09-981-353-34

Query Match 4.3%; Score 115.4; DB 10; Length 2574;
 Best Local Similarity 51.2%; Pred. No. 6.5e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

548 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGCTTGGCCCTCA 607
 776 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGAGCTTGGCCCTCA 835
 608 TTGCGCAGATGGCGGCTGCCAGAACCTTCTCCAGAGCCAGGCTACAAAAGGTGAGG 667

836 CAGCAGCTCTTGGGCTTGCACAACTATGAGAGATGAGAAAAAACGACGAGATTACA 895
 668 ACCTATGCGGAGCTGGGCTTGAAGCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 727
 896 GGGTCATTCAGAGATGAGTCTGATTAAGTGGAGACTCCAAAGGTGAACTCAGTTTA 955
 728 TCGCGGGGTGTGCGGGGGGTGAGCGCAGAGAGATGACATTTGGGCTGACGCTCTGGA 787
 956 TCCGAGGTGTCTGAG 1015
 788 ACCAGGAATCTTATTCGACAGCAAGACCCAGCTTGGCGCTGACAGCTTACAGCCACA 847
 1016 ATCTTCATCTTGTCTTGTGATGAGCTTACAACTGCGCTTGAAGCTCAAGACAGCAATG 1075
 848 ACCTGATGACAGCTTGTGACAGCAAGACCCAGCTTGGCGCTGACAGCTTACAGCTCC 907
 1076 CTGCTCTTCTCTCTGAAAAGAGATGTTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
 908 ACCAGCTGCGCTGACATCTTCAAGCTGCTTGAATCTGCTCTCTGATGACGTGGCA 967
 1136 ATCAGCTGAGATCTTCAATCTTCAAGTGTGATGAGCTTACCTTATTTGGCCTCAGAGAA 1195
 968 CCCCCATCTACTAGGAGGCGGCGCCAGACATGCTCAGATTTTCACAGCCATCGGCTACC 1027
 1196 GACTTATGTTCCAGAGGCGCTCTCAGAGAGGCGCTTGGATTTGATGATCAGCTGTTATC 1255
 1028 CCTGCTCTGCTTACAGCAATCTGCTGATCTTATGAGACCTGA 1072
 1256 ACTGTGAGGCTTATATATACCTGACACTTCTTCTGACATCA 1300

RESULT 11

US-10-120-687-60
 : Sequence 60, Application US/10120687
 : Publication No. US20030082155A1
 : GENERAL INFORMATION:
 : APPLICANT: Massachusetts General Hospital
 : TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treat
 : FILE REFERENCE: 3284/1235B
 : CURRENT APPLICATION NUMBER: US/10/120.687
 : CURRENT FILING DATE: 2002-04-11
 : PRIOR APPLICATION NUMBER: US60/169082
 : PRIOR FILING DATE: 1999-12-06
 : PRIOR APPLICATION NUMBER: US 09/963,875
 : PRIOR FILING DATE: 2001-09-25
 : PRIOR APPLICATION NUMBER: US 60/215109
 : PRIOR FILING DATE: 2000-06-28
 : PRIOR APPLICATION NUMBER: US 60/238880
 : PRIOR FILING DATE: 2000-10-06
 : PRIOR APPLICATION NUMBER: US 09/731261
 : PRIOR FILING DATE: 2000-12-06
 : NUMBER OF SEQ ID NOS: 61
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 60
 : LENGTH: 2718
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 US-10-120-687-60

Query Match 4.3%; Score 115.4; DB 14; Length 2718;
 Best Local Similarity 51.2%; Pred. No. 6.6e-24;
 Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

548 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGCTTGGCCCTCA 607
 572 ACCTGCGCCAGACACACAGCTGCTCCCAACTGACGTGGAGAGAGAGCTTGGCCCTCA 631
 608 TTGCGCAGATGGCGGCTGCCAGAACCTTCTCCAGAGCCAGGCTGACAAAAGGTGAGG 667
 632 CAGCAGCTCTTGGGCTTGCACAACTATGAGAGATGAGAAAAAACGACGAGATTACA 691

Db 909 TCAGTGTGAGCGCAAGCGCTGGCCATGCGCGTGGAGCTGGTGAACAACCTCCAGTC 968
QY 799 CTATTTTCAGCAACCCACCTCTGGCTGAGACGTTCAACGCCACAACCTGTGTAG 858
Db 969 ATGTTCTTCATATGAGCCACCGCTGAGAGCGCTCTGTTCCAGGTGCTCG 1028
QY 859 ACCTTGTCCAGGCTGGCCAAAGCAACCGCTGTGTCTATCTCCCTCCACCGCTCG 918
Db 1029 CTATGAAAGGCTCGCTCAAGGGGCTGCTCATATTTTGCACATCCACCGCCAGC 1008
QY 919 TCTGACATCTTACGCTGTTTGAATGCTCTCTGATGAGCTGCTGACCCCATCTAC 978
Db 1089 GCCAACTCTTGAGCTGTTCAGACAGCTTTAGCTCTGATGATGAGCAATGTGTAC 1148
QY 979 TTAGGGGGCCCGCAGCAATGCTGCAATTTTCAAGCCATCGGCTACCTGCTCG 1038
Db 1149 CGGGGAAAGTCTGCAATCTTGTGCAATTTTGAAGGATTTGGGTTGAACTGCCAAC 1208
QY 1039 TACAGCATCTGCTGCTGCTTCTATGAGCT---GACCAGATTTACAGGGCCACAGA 1095
Db 1209 TACACACACCCAGCAGATTTTGTATGAGGTTGCAATCGGCGAGTACGTTATCAAG 1268
QY 1096 GAGCAGGAATTTGGCCACAGGGAGAGGCTCACTGCACTGCAAGCCCTGTTTGAAGAA 1155
Db 1269 AGTGGCTGTGAGAGCGGTTGGGAGGGCATGTGACTGACACCAAGAGAGCTTC 1328
QY 1156 GTGCGCTATGATGATCTTCTATGAAAGACAGAGAGATTTGACGAGGACCC 1215
Db 1329 GGGGGTATGCGGAGGTGAACCTTTCTTGGCACCGGCTCTGACAGGATTAAGCAG 1388
QY 1216 TGTGTGAAGAGAGGTGAGCCCACTAGACCACTGCTCCCGAGTCTTACAGAGATG 1275
Db 1389 ACMAAGATTTAAAGGGTTGAAGAAAGACTCTCTCATGAAAGGCGCCACAGCTTC 1448
QY 1276 CCTG---GGGCGGTGAGCAGTTTACGAGCTGATCCGTGATTTTCAACGACTTC 1332
Db 1449 TCGCAGCTGTGCTACGAGATTTTGCATCTTCAAGAGACCTTCTCAGCATCAG 1508
QY 1333 CGAGACCTGCGCAACCTCTCTCATCTGAGGGGCGAGGCTGTGTATGTCATGACATC 1392
Db 1509 AAGGAGCTGCTCTGACACACCTGCGCATCACTGCGCATTTGGCATTGGCTCGCTCAT 1568
QY 1393 GCTTCTCTATTTTGGCCATGAGGAGCATCCAGCTCTCTCATGATGATACAGCCCTTC 1452
Db 1569 GCGCTGCTTACTTGGGATCGGGAACGAAAGGCTTTGAGCAACTCCGCTTC 1628
QY 1453 TTGTCATGATGCTGCTCTCATCTCTTCAACGTCATTTGATGTCATCTCAAAATG 1512
Db 1629 CTCTCTCTCATGCTGTCTCTCAAGTTGCGGGCCCTCATGCTACTGTTTGTACATTT 1688
QY 1513 TACTAGAGAGGCGCAATGCTTACTATGAACTGGAAGAGGCTGTATACACACTGCTCA 1572
Db 1689 CCGCTGAGATGGAGTCTTCTTCCGGGAACACTGAACCTACTGCTAGCTGAAAGCC 1748
QY 1573 TATTTCTTGGCAATCTCTGGGAGGCTCCGGAGCACTGTGCTTACATCTCATCTAC 1632
Db 1749 TACTATCTGCGCAACATGAGCAACGCTGCTTCAAGATCTTCCAGTGGCTAC 1808
QY 1633 GGGATGCCACCTACTGCTGCGCAACCTGAGGCGCTTCAAGCTCTCTCTGCTGAC 1692
Db 1809 TGCACATCTGCTACTGATGAGCTGCGAGCGCTCGAGCGCGCTGCTGTGTGTTT 1868
QY 1693 TTTCTGCTGTGCTGCTGTGTCTTCTTGTGAGAGATTTAGGCGCTGCGCGCGGCTG 1752
Db 1869 GCGCGGCTGGAGCACTCTCTCTGTGCGACAGTCCCTGCTGCTGATGAGAGCC 1928
QY 1753 CTGCTCCCACTTCTCCATGAGCTCTTCTTCAACATGGCCCTTACAGCTCTTAC 1812
Db 1929 GCGTCCAGCTCTCTGAGTGGCCACTTCTGCGCCAGTGAACGCCATCCCGGTCTC 1988
QY 1813 CTGCGCGGCTCTCATATAAACTTGAAGAGCTGTGAGACAGTCCCGCGGTGATTTCC 1872
Db 1989 CTGTTCTGCGGGTCTCTGCTGAGCTTGCAGACACATCCCACTACTACAGTGAATGCTC 2048

QY 1873 AAGTGTCTCTCGCGGTGTGTGTTTGAAGGCTCAT 1910
Db 2049 TACATCTCTATGTGATGATGCTTCAAGAGGCTCAT 2086

RESULT 14

US-09-954-531-591
Sequence 591, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
FILE OF INVENTION: Gene Sets
FILE REFERENCE: 689290-77
CURRENT FILING DATE: US/09/954,531
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 591
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-591

Query Match 3.9%; Score 105.2; DB 10; Length 2930;
Best Local Similarity 44.3%; Pred. No. 8.4e-21;
Matches 682; Conservative 0; Mismatches 823; Indels 33; Gaps 5;

QY 379 AAGTGAAGAGGGGCAATGCTGCTCATATGAGGAGCTCAAGTTGTGGAGACCTTC 438
Db 337 AAGTTCAATGAGTGTGATGTTGTGTGCTTATGAGGCTTCCGGGCGGGAAGTCCAG 396
QY 439 TTGCTAGATGTGATCTGCGGAGGTACAGCGGCGCAAGATCAATGATCAAGCCATGTGG 498
Db 397 CTGATGAACATCTGCTGCT---GGATACAGAGGAGAGGGCATTAAGGGGCGCTCTC 450
QY 499 ATCAATGCGCAGCCAGCTCGCTGCTGAGTGTGAGAAAGTGTGAGGCGGCGGCGAG 558
Db 451 ATCAAGGCGCTGCGGCGGAGCTGCTGCTTCCGGAAGGTGTCTGCTACATATGACAG 510
QY 559 CACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618
Db 511 GATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570
QY 619 CGGCTGCCAGAACCTTCTCCAGGCGCGGCTGCAACAAAGGAGTGTGAGAGCTGTGCGG 678
Db 571 AAGCTTCAAGGAAGG-----ATGAAGCAGAAAGGAAATGTCTCAAGAGATACTGACA 624
QY 679 GAGCTGCGGCTTAGCAATGCTGCTGACACCGCGCTGAGCAACATGATGCTGCGGGGCTG 738
Db 625 GCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 669
QY 739 TCGGGGGGTGAGCGAGAGAGTCAAGTGGGGTGGAGCTGCTGCTGCTGCTGCTGCTGCTG 798
Db 670 TCAGTGTGATGAGCGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
QY 799 CTATTTCTGAGCAACCCACTTGTGGCTGTGAGCTTCAAGAGCCCAACCTGCTGTAG 858
Db 730 ATGTTCTTCAATGAGCCACAGCGGCTGGAAGCGGCTCTCTCTTCAAGTGTGCTCG 789
QY 859 ACCTTGTCCAGCTGCGCAAGGCAACCGGCTGTGCTCATCTCTTCCACCACTGCTCG 918

Db 790 CTGATGAAGGCGCTCGCTCAAGGGGCTCCATCATATTCACCAATCCACCCAGC 849
OY 919 TGTGACATCTTCAGGCTGTTTGAATGTCCTCCCTGATGACGTCGACCCCATCTAC 978
Db 850 GCCAAACTCTTCGAGCTTCGACACACTTACCTCTGATGACAGCAATGCTGTAC 909
OY 979 TTAGGGGCGCCGACGACATGCTCAGTATTTACAGCCATCGGCTACCCCTGCTCCG 1038
Db 910 CGGGGAAAGCTTCGATCTTGTGCGCATATTTGAGGATTTGGGTGTGAAGTCCCAAC 969
OY 1039 TACAGCAATCCTGCTGCTCTATGAGGACCT---GACCAGCATGACAGCCAGAGA 1095
Db 970 TACCAACACCCAGACGATTTTGTATGAGAGGTCATCGGCGAGTACGGTGTATCAAC 1029
OY 1096 GACAGCAATTTGGCCACACGAGGAGAGGCTCATGCTCAGCCCTGTTTCAAGAAA 1155
Db 1030 AGTCGGCTGTGAGAGGGGTTGCGAGGGCATGTGTACCTACACACCAAGAGAGACCTC 1089
OY 1156 GTGCGTGTACTTATGATGACTTCTTATGAGAAAGAGAGAGAGATCTTGAAGAGACC 1215
Db 1090 GGGGTGTATGCCAGGTGAACCTTTCTTTGGCACCGCCCTCTGGAAGGTAAAGCAG 1149
OY 1216 TGTGTGAAAGAGCGGTGACCCCACTAGACCACTGCTCCGAGTCTTACAGAGATG 1275
Db 1150 ACAAAGATTAAGGGGTTGAGAAAGAGTCTCTGTCATGGAAGGCTCCACAGCTTC 1209
OY 1276 CCTG---GGGCGGTGACAGATTACGACGCTGATCCGTCGATGATTTCCAGAGACTTC 1332
Db 1210 TGTGACAGTGGCTCAGGAGCTTCTGATCTCTTCAAGAGACCTTCTCAGACATCAG 1269
OY 1333 CGAGACCTTGGCCACCTCTCATATCAGGGGCGAGGCTGTGTATGTCATGATGACATC 1392
Db 1270 AGGACCTGGTCTCTGACACCTGCGCATATCCTGCAATTTGGGATCGGCTCTGATC 1329
OY 1393 GGGTCTCTATTTTGGGATGAGGAGCATTCAGCTCTCTTCAATGATACAGCCGCTTC 1452
Db 1330 GGGCTCTGACTTGGGAGTGGGAAAGAAACCAAGAGTCTTGAAGCACTCGGCTTC 1389
OY 1453 TTGTCATATGCTGCTCTCATATCCCTTCAACGTCATCTGATGATGATCTCAATATG 1512
Db 1390 CTTCTTCTCCATGCTGTTTCCATGTTTGGGAGGCTTATGCTTCTGATGATATTT 1449
OY 1513 TACTAGAGAGGCAATGCTTATGATGAACTGGAAGAGGCGCTGTACACCACTGGTCA 1572
Db 1450 CCCCTGAGATGGAGTCTTCTTGGGAAACCTGATACGTCGATACAGCTGGAAGGCC 1509
OY 1573 TATTTCTTGGCAAGTCTCGGGAGCTTCGAGACCTGTGCTTACATCATCATCTAC 1632
Db 1510 TACTACTGCGCCAAAGCAATGAGAGAGGCTTCCATGATCTTCCAGTGGGCTAC 1569
OY 1633 GGGATCCCACTACTGCTGCGGCAACGTCGAGGCGAGGCTCCAGGCTCTCTGCTGCAAC 1692
Db 1570 TGCAGATGCTGATGAGATGAGTGCAGCCGCTCGAGCGGCTTGTGCTGCTT 1629
OY 1693 TTCTGCTGCTGTGCTGTGTGCTTCTGTCAGAGATTAAGCCCTGCGCGCGGCGC 1752
Db 1630 GCGGCGCTGGGACATGACCTCCCTGAGGCAACATCCCTGGGCTGCTGATCGGAGCC 1689
OY 1753 CTGCTCCCACTTCCACATGCGCTCTCTTCTGAGCAATGCCCTCTACAACTCTTCTAC 1812
Db 1690 GCGTCCAGCTCCCTGAGGTGCGCACTTCTGTCGGGCGCAGTACAGACCACTCCGAGCTC 1749
OY 1813 CTGCGCGGGGCTTCAATGATTAACCTGACAGCTGTGACAGTGGCGCGGAGATTCG 1872
Db 1750 CTGTTCTGGGGGTTCTTCTCACTTCTGACACCAATCCCACTTACAGTGTGATCC 1809
OY 1873 AAAGTCTCTCTGCTGTGCTGCTTCTTGAAGGCTGAT 1910
Db 1810 TACATCTCTATGTAGGTATGGGTTTGAAGGGGCTAT 1847

RESULT 15
US-10-171-581-276

Sequence 276, Application US/10171581
Publication No. US20030104426A1
GENERAL INFORMATION:
APPLICANT: Dai, Hongyue
APPLICANT: Linsley, Peter
APPLICANT: Mao Mao
TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
FILE REFERENCE: 9301-157-999
CURRENT APPLICATION NUMBER: US/10/171,581
CURRENT FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: 60/298,914
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 366
SEQ ID NO 276
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
PUBLICATION ACCESSION NUMBER: X91249
DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-276

Query Match 3.98; Score 105.2; DB 14; Length 2930;
Best Local Similarity 44.3%; Pred. No. 8.4e-21;
Matches 682; Conservative 0; Mismatches 823; Indels 33; Gaps 5;

OY 379 AAAGTGAAGTGGGAGATGCTGGCCATCATAGGAGCTCAGGTTGTGGAGACCTCC 438
Db 337 AAGTTCAATAGTGTAGTGTGAGTGTGCTTCCCGGGCCGGGAGATCCAGC 396
OY 439 TTGCTAGATGATCTACCTGCGCGGAGTCAAGGAGTCAAGTCAAGGAGGCTGCG 498
Db 397 CTGATACATCTGCTGCT-----GGATACAGGAGACGGGATGAGAGGGGCGCTTC 450
OY 499 ATCAATGAGGAGCCGACGCTCGCTCAGCTGTGTGAGGAGTGTGTGCGCCACGTCGCCAG 558
Db 451 ATCAAGGCGCTGCGCGGAGACCTGCGCTGCTTCCGGAAGGTGCTGCTACATCATGACG 510
OY 559 CACACACCTGCTCCCACTTACTGTGCGAGAGACCTTGGCTTCAATGCGCCAGATG 618
Db 511 GATGATCATGCTGTGCGGATCTCACTGTGACAGAGCCATGATGTGTGCGCCATCTG 570
OY 619 CGGCTGCCAGAACTTCTCCAGGCGCCAGCGTGAAGAAAGGTGAGAGAGCTGATCGG 678
Db 571 AAGCTCAGAGAGAG-----ATGAAGGCAAGAGGAAATGTCTAAGGAGATATGACA 624
OY 679 GAGCTGCGGCTTAGGAGTGCCTGACACCCGCGTGGGCAACATGATGCTGCGGGGTTG 738
Db 625 GCGCTGCGCTGTGCTGCTGCGGCAACAGCGGAGCG-----GAGCTG 669
OY 739 TCGGGGGTGAAGGAGAGAGTACAGATTTGGGTGACGCTCTGTGAACCCAGGATC 798
Db 670 TTAGGTGTACAGGAGAGCGGCTGCGCATCGGCTGAGTGTGTAACACCTCCAGTC 729
OY 799 CTATTTCTGAGAGAACCCACTCTGCGCTGAGAGCTTACAGACCCCAACCTGAGAG 858
Db 730 ATGTTCTTGAAGAGCCACAGAGGCTGAGAGGCTCTCTGCTTCAAGTGTGCTG 789
OY 859 ACCCTGTCAGGCTGCGCAAGGCAACCGGCTGTGCTCATATCTCCCTCAGCAGCTGCG 918
Db 790 CTGATGAAGAGGCTGCTCAAGGGGTGCTCATCATATTTGACACCATCCACAGCCGAC 849
OY 919 TGTGATCTTGAAGCTGTTTATGCTGTCTCTGATGACGTCGACACCCCATCTAC 978
Db 850 GCCAAACTCTTGAAGTGTTCAGCAGCTTACGTCCTGATGAGCAAGCAATGTGTGTC 909
OY 979 TTAGGGGCGCCAGACATGCTGATTTTCAACAGCATGAGGCTACCCCTGTCTCC 1038
Db 910 CGGGGAAAGTGTCAATCTTGTGCCATATTTGAGGAGATTTGGTGTGAAGTCCCAAC 969
OY 1039 TACAGCAATCCTGCTGCTCTATGAGGACCT---GACCAGATTGACAGGCGGACAGACA 1095
Db 970 TACCAACACCCAGACGATTTTGTATGAGGTTGATCCGCGGAGTACGGTATCAGAAC 1029

Search completed: July 27, 2003, 08:56:18
Job time : 583.074 secs

OY 1545 GGAAGACGGCTGTACACCACTGGTCATATTTCTTCCAGATCCTGGGAGCTTCC 1604
DB 1184 YY 1243
OY 1605 GGACACCTGCTCATATCATTACATGAGATGCCACCTACTGGCTGGCAACTGAG 1664
DB 1244 YY 1303
OY 1665 GCCAGGCTCCAGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1724
DB 1304 YY 1363
OY 1725 CAGGATATATGCGCTGGCGCGCGCGCTGCTGCCACCTCCACATGCGCTCTT 1784
DB 1364 YY 1423
OY 1785 CAGCAATGCCCTTACAACTCTCTTACTCTC 1815
DB 1424 YYYYYYYYYYGTACCAATCTCTTACTCTC 1454

RESULT 8

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match

Best Local Similarity 49.08; Score 54; DB 3; Length 4403765;

Matches 187; Conservative 0; Mismatches 180; Indels 15; Gaps 1;

OY 528 GGTGAGAGAGTGTGCGCCACGTGCGCCAGACACACACCTGCTCCCACTTACTGT 587
DB 1965623 GCTGCGAGCAGAGATGCGGATGCGCCAGAGAGAGAGAGTGTGACGCGT 1965682
OY 588 GCGAGAGACCTTGCTTCAATGCGGATGCGGCTGCCAGAACCTTTCGCCAGGCCA 647
DB 1965683 GAAACACGCGCTGATGTATGCGCGGAGACTACGAGCTGCCCGGACACCAAAAGATGA 1965742
OY 648 GCGTACAAAAGAGTGTGAGAGACGTATGCGGAGCTCGGCTTATGAGAGTGGCTGAC 707
DB 1965743 CCGGACCCAGGATGTTCCCGGGGTCTCGAAGAACTGGAATGTCCAAAGACATCGACAC 1965802
OY 708 CCGGCTGGGCAACATGTACGTGCGGGGTGCGGGGAGCGCAGAGATCAGCAT 767
DB 1965803 CAGGTGCGACAA-----GCTGTCGGGTGTCAGACGCAAGCGGGGTGCGGT 1965847
OY 768 TGGGGTGCAGCTCTGTGGAACCCAGAAATCTTATCTGTGAGAGAACCCACTCTGGCT 827
DB 1965848 GGGCTTGAACCTGTGACCGGGGCGGTACCTGTATCTGTGAGAGCGGACATCCGGCT 1965907
OY 828 CGACAGCTTACAGCCCAACCTGTGAGAGCTTGTCCAGGCTGGCCAAAGGCAACCG 887
DB CGACAGCTTACAGCCCAACCTGTGAGAGCTTGTCCAGGCTGGCCAAAGGCAACCG 887

DB 1965908 AGATCTGCGCTGAGACCGGAGGTATGATGATGCTGCGGAGATTGGCCAGCGGGTGC 196596.
OY 888 GCTGTGCTATCTCCCTCCAC 909
DB 1965968 GGTGTGCTCTGCTGTTACCCAC 1965989

RESULT 9

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328

GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match

Best Local Similarity 49.08; Score 54; DB 3; Length 4411529;

Matches 187; Conservative 0; Mismatches 180; Indels 15; Gaps 1;

OY 528 GGTGAGAGAGTGTGCGCCACGTGCGCCAGACACACACCTGCTCCCACTTACTGT 587
DB 1974794 GCTGCGAGCAGAGATGCGGATGCGCCAGAGAGAGAGTGTGACGCGT 1974853
OY 588 GCGAGAGACCTTGCTTCAATGCGGATGCGGCTGCCAGAACCTTTCGCCAGGCCA 647
DB 1974854 GAAACACGCGCTGATGTATGCGCGGAGACTACGAGCTGCCCGGACACCAAAAGATGA 1974913
OY 648 GCGTACAAAAGAGTGTGAGAGACGTATGCGGAGCTCGGCTTATGAGAGTGGCTGAC 707
DB 1974914 CCGGACCCAGGATGTTCCCGGGGTCTCGAAGAACTGGAATGTCCAAAGACATCGACAC 1974973
OY 708 CCGGCTGGGCAACATGTACGTGCGGGGTGCGGGGAGCGCAGAGATCAGCAT 767
DB 1974974 CAGGTGCGACAA-----GCTGTCGGGTGTCAGACGCAAGCGGGGTGCGGT 1975018
OY 768 TGGGGTGCAGCTCTGTGGAACCCAGAAATCTTATCTGTGAGAGAACCCACTCTGGCT 827
DB 1975019 GGGCTTGAACCTGTGACCGGGGCGGTACCTGTATCTGTGAGAGCGGACATCCGGCT 1975078
OY 828 CGACAGCTTACAGCCCAACCTGTGAGAGCTTGTCCAGGCTGGCCAAAGGCAACCG 887
DB 1975079 AGATCTGCGCTGAGACCGGAGGTATGATGATGCTGCGGAGATTGGCCAGCGGGTGC 1975138
OY 888 GCTGTGCTATCTCCCTCCAC 909
DB 1975139 GGTGTGCTCTGCTGTTACCCAC 1975160

RESULT 10

US-09-252-991A-11541
Sequence 11541, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMO

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

Query Match	1.9%;	Score 51.4;	DB 4;	length 723;
Best Local Similarity	46.0%;	Pred. No. 0.0025;		
Matches 212;	Conservative	0;	Mismatches 246;	Indels 3;
				Gaps 1;

RESULT 11
US-09-252-991A-11845/C
: Sequence 11845, Application US/09252991A
: Patent No. 6551706

Query Match	1.98;	Score 51.4;	DB 4;	Length 1155;
Best Local Similarity	46.08;	Pred. No. 0.0029;		

QY	468	CGGGGGGCAAGTCAAGTCAAGGCGAGATCTGGATCAATGGGACGCCAGCTGGCCTCACT	527
Db	859	CGCCGGGCAAGACACACCACTATCAACTGTGCTCTGGCCCTGCTGGGCCCCAGCAAGGCCG	800
QY	528	GGTGAAGAAGTGTGTGGCCCAAGTCGCGCAGACACAAACAGCTGGCTCCCCAATTGACTGT	587
Db	799	CGTGGGGTCCCTCGGGCCCAAGATGGCAGGAGGCTTGGAGGGCGCGCCGCAACTTCGGCTACCT	740
QY	588	GGGAGAGACTTTGGCCCTTCATTGGCCAGATGGCGGCTGGCCAGAACCTTCTCCAGGCCCA	647
Db	739	GCCGGAGAGACTGACCTTCTACCCGCACTCAGGGCGCGGAGAAACCTCGCCCACTTTCG	680
QY	648	GGGTGACAAAAGGGTGGAGGAGCTGATGGCGAGCTGGGGTTGGAGCGAGTGGCGTGAAC	707
Db	679	CGGCTCAAGGGGCGTGGGGCGGGCGCGGAGACCGCGCGCCTGCTGGAAACAGTTCGGCTGG	620
QY	708	CCGCGTGGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAAGCGCAGAGAGTCAACAT	767
Db	619	CCATGCACCCACAGGGGGCGCCTGTAACCACTTACTCAAGGGGCAATGGCCAGCGCTCGGGCT	560
QY	768	TGGGGTGAAGCTCGTGTGGAGCCAGGAAATCTTATTCGTGACGACACCCACTCTGGGCT	827
Db	559	GGCCGAGGCGCTGCTCGGCGAACCAGCGCTCTGTGCTGCACGAAACGAGGTGGGCT	500
QY	828	CGACAGCTTCAACAGCCCAACAACCTGGTGAAGACCTGTCCAGCGTG---GCCAAGGCA	884
Db	499	CGACCCGCTGGCCACCGTCGAGCTCTACCAATTGCTGCAGCGGCTGGCGGGCCAGGGGAC	440
QY	885	CCGGGTGGTCAATTCCTCCACACCGAGCTTCGCTGACA	925
Db	439	CGGATGCTCTTTGCTCCCAATGATGCTGCGCGGGTGA	399

Query Match	1.9%	Score 51.4;	DB 4;	Length 2367;
Best Local Similarity	46.0%;	Pred. No. 0.0038;		
Matches 212; Conservative	0;	Mismatches 246;	Indels 3;	Gaps 1.

[illegible]

OY 648 GCTGTACAAAAGGTTGAGAGAGTATCGGAGAGCTGGGCTTACGAGTGGGCTGACAC 707
DB 1749 CCGGCTCAAGGGGCGTGGGGCGGCGGAGAGCGGCGCTGCTGGAGACAGTGGGCTGG 1808
OY 708 CCGGCTGGGCAACATGATGTCGGGGGTTGTGGGGGGTGGAGCGGAGAGTGCAGCAT 767
DB 1809 CCATGTACAGCCAGGCGGCGCTGAAAACCTACTCGAAGGGCATGCGCCAGCGGCTGGCT 1868
OY 768 TGGGGTGCAGCTCTGTGTGGAACCCAGAACTTATTTTCAGACCAACCCTCTGGCT 827
DB 1869 GCGCCAGGCGCTGTGCTGGCGAAGCGGCTGCTGTGTGTGACGAGACGAGTGGGCT 1928
OY 828 CGACAGCTTCACAGCCCAACATGTCGTAAGACCTTGTCCAGGCTG---GGCAAGGCA 884
DB 1929 CGACCGGCTGGCGACCGCTGAGAGCTTACCAATGCTGACAGCGCTGCGGCGGCGAC 1988
OY 885 CCGGCTGGTGTCTATCTCCCTCCACAGCGCTGCTGACA 925
DB 1989 CGGATCGCTCTTGTCTCCCATGTGTGTCGCCGCGGTGAGA 2029

RESULT 13

US-08-592-874-1/c
Sequence 1, Application US/08592874
Patent No. 5854034

GENERAL INFORMATION:

APPLICANT: POLLOCK, THOMAS J.
APPLICANT: YAMAZAKI, MOTOHIDE
APPLICANT: THORNE, LINDA
APPLICANT: MIKOLAJCZAK, MARCIA
APPLICANT: ARMENTROUT, RICHARD W.
TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: JULES E. GOLDBERG
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016-2391

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,874
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/377,440
FILING DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 28804 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
US-08-592-874-1

Query Match 1.9%; Score 51.4; DB 2; Length 28804;
Best Local Similarity 47.9%; Pred. No. 0.0094;
Matches 148; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

OY 606 CATTGCCAGATGGGGTGTGCCAGAACCTTCTCCAGGCCCAAGCTGACAAAGGTGGA 665
DB 19351 CATTGCCGTGTCCAAACCCGGGAGATGCCCTGTGAGATGTCGTGGGGGAGACCTGGC 19292
OY 666 GGAGCTGATTCGAGAGTGGCGCTTAGGAGTGGCGTACACCCGCGTGGGCAACATGTA 725
DB 19291 GGGTGGCATATCACTTATCTGCTGACCGCGGCTGATGACACCGAGATGTCGACCG 19232
OY 726 CGTGGGGGGTGTGGGGGGTGAACCGAGAGAGTACGATTTGGGTGTCAGCTCTGTG 785
DB 19231 CGGGGTCAACCTGTGTGGGGGCGACCGCGAGGGGCTGCTATGCGCCGGGCTGTGG 19172
OY 786 GAACCCAGAAATCTTATTTTCGACGAAACCACTTGTGGCTGACAGCTTCACAGCCCA 845
DB 19171 CAATCCGGCATCTGTGTGTTCGACGAGGCACTCCGCTGTGATGCCGAGAGCGAGA 19112
OY 846 CAACCTGTGTAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGCGTGGTCAATCTCCCT 905
DB 19111 GCTGATCCAGAACATCTGCGCGCATCTCGGGGGCGCCACGCTGTGATCATGCGCCA 19052
OY 906 CCACCGAGCC 914
DB 19051 CCGGCTGTC 19043

RESULT 14

US-09-096-942-2/c
Sequence 2, Application US/09096942
Patent No. 6027925

GENERAL INFORMATION:

APPLICANT: Pollock, Thomas J.
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
APPLICANT: ArmentROUT, Richard W.
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT FILING DATE: 1998-06-12
EARLIER APPLICATION NUMBER: US/09/096,942
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 28804
TYPE: DNA
ORGANISM: Sphingomonas sp. 588
US-09-096-942-2

Query Match 1.9%; Score 51.4; DB 3; Length 28804;
Best Local Similarity 47.9%; Pred. No. 0.0094;
Matches 148; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

OY 606 CATTGCCAGATGGGGTGTGCCAGAACCTTCTCCAGGCCCAAGCTGACAAAGGTGGA 665
DB 19351 CATTGCCGTGTCCAAACCCGGGAGATGCCCTGTGAGATGTCGTGGGGGAGACCTGGC 19292
OY 666 GGAGCTGATTCGAGAGTGGCGCTTAGGAGTGGCGTACACCCGCGTGGGCAACATGTA 725
DB 19291 GGGTGGCATATCACTTATCTGCTGACCGCGGCTGATGACACCGAGATGTCGACCG 19232
OY 726 CGTGGGGGGTGTGGGGGGTGAACCGAGAGAGTACGATTTGGGTGTCAGCTCTGTG 785
DB 19231 CGGGGTCAACCTGTGTGGGGGCGACCGCGAGGGGCTGCTATGCGCCGGGCTGTGG 19172
OY 786 GAACCCAGAAATCTTATTTTCGACGAAACCACTTGTGGCTGACAGCTTCACAGCCCA 845
DB 19171 CAATCCGGCATCTGTGTGTTCGACGAGGCACTCCGCTGTGATGCCGAGAGCGAGA 19112
OY 846 CAACCTGTGTAAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGTGTGCTCATCTCCCT 905
DB 19111 GCTGATCCAGAACATCTGCGCGCATCTCGGGGGCGCCACGCTGTGATCATGCGCCA 19052

DY	906	CCACCAGCC	914
Db	19051	CCGCGTGTC	19043

RESULT 15

```

US-09-096-867-2/c
Sequence 2, Application US/09096867
Patent No. 6030817
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
APPLICANT: Amenttrott, Richard W
TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,867
EARLIER FILING DATE: 1998-06-11
EARLIER APPLICATION NUMBER: 60/049,428
EARLIER FILING DATE: 1997-06-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentlin Ver. 2.0
SEQ ID NO 2
LENGTH: 28804
TYPE: DNA
ORGANISM: Sphingomonas sp. 588
US-09-096-867-2

```

Query Match	1.93	Score	51.4	DB	3	Length	28804
Best Local Similarity	47.98	Pred. No.	0.0094				
Matches	148	Conservative	0	Mismatches	161	Indels	0
						Gaps	0

QY	606	CATTGCCAATGGGTGGCCCGAAGACCTTCCTCCAGGCCCAAGCTGACAAAGGGTGA	665
Db	19351	CATTGGCCTGCCAACCGCGATGCCGTTGAGCAATGCTGGCGGGCAGCCTGGCC	192922
QY	666	GGAAGTAAATCGCGAGCTGGCGCTTAGCCAGTAGTCGCTGACACCCGGTGGGCAATGTA	725
Db	19291	GGGTGGCGCATGACTTCATCTCCGTGAGCGCCCGGCTGATGACACCGAATCGTGAGGCG	192322
QY	726	CGTGGCGGGTGTCTGGCGGGGTGAGCGGAGAGACATTTGGGGTGCACCTCTGTG	785
Db	19231	CGGGCTGACACTTCCTGGGGCGGCCAGCGCCACGCGCTGCTATGCGCCGCGCCTGTGCG	191722
QY	786	GAAACCCAGAACTCTTAATCTCGACGAAACCAACCTCTGGCTCGAAGCTTGACAGCCCA	845
Db	19171	CAATCGCGGCACTCTGTGTTCCACGACGAGCGACCTCCGGCGTGGAATGCCGAGAGGAGA	191122
QY	846	CAACTGGTGAACCTTGTCGACGGCTGGCGCAAAAGGCACACGGGTGGTCAATCTCCCT	905
Db	19111	GCATGTCAGAAACAAATCTGCGGCCCATCTTGGCGGGCGCAGCGTGGTCAATCATGCCCA	190522
QY	906	CCACGACGCC	914
Db	19051	CCGCGCTGTC	19043

Search completed: July 27, 2003, 12:45:54
Job time : 182.378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:04:42 ; Search time 15 Seconds

(without alignments)
1800.073 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MCDLSSLTFCGSMGLQVNRG.....PALVIGIVKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3326	100.0	651	1 ABG5_HUMAN	Q9H222 homo sapien
2	2738.5	82.3	652	1 ABG5_MOUSE	Q99pe8 mus musculu
3	2727.5	82.0	652	1 ABG5_RAT	Q99pe7 rattus norv
4	698	21.0	673	1 ABG8_MOUSE	Q9d8m0 mus musculu
5	697	21.0	673	1 ABG8_HUMAN	Q9H221 homo sapien
6	691	20.8	672	1 ABG8_RAT	P58428 rattus norv
7	676.5	20.3	655	1 ABG2_HUMAN	Q9unq0 homo sapien
8	627	18.9	1294	1 YOH5_YEAST	Q08234 saccharomyc
9	623	18.7	677	1 WHIT_LOCCU	Q05360 lucilia cup
10	621	18.7	1049	1 ADP1_YEAST	P25371 anopheles a
11	607.5	18.3	695	1 WHIT_ANOGA	Q27256 anopheles g
12	602.5	18.1	687	1 WHIT_DROME	P10090 drosophila
13	596.5	17.9	678	1 ABG1_HUMAN	P45844 homo sapien
14	591	17.8	679	1 WHIT_CERCA	Q17320 ceratitlis c
15	578.5	17.4	646	1 ABG4_HUMAN	Q9H172 homo sapien
16	573	17.2	709	1 WHIT_ANOAL	Q16928 anopheles a
17	569.5	17.1	666	1 ABG1_MOUSE	Q64343 mus musculu
18	561.5	16.9	598	1 YPC3_CAEEL	Q11180 caenorhabdi
19	547	16.4	666	1 SCRT_DROME	P45843 caenorhabdi
20	517.5	15.6	610	1 YOS_CAEEL	Q09466 caenorhabdi
21	454.5	13.7	675	1 BROW_DROME	P12428 drosophila
22	435	13.1	668	1 BROW_DROME	Q24739 drosophila
23	427	12.8	1499	1 CDR2_CANAL	P78595 candida alb
24	406.5	12.2	1333	1 YN99_YEAST	P53756 saccharomyc
25	403.5	12.1	1364	1 YN99_YEAST	P53756 saccharomyc
26	399.5	12.0	1530	1 BFR1_SCHPO	P41820 schizosacch
27	398	12.0	1501	1 SNQ2_YEAST	P32568 saccharomyc
28	392.5	11.8	1501	1 CDR3_CANAL	Q42890 candida alb
29	391	11.8	1529	1 PDRF_YEAST	Q04182 saccharomyc
30	383.5	11.5	650	1 ABG3_MOUSE	Q99p81 mus musculu
31	382	11.5	1501	1 CDR1_CANAL	P43071 candida alb
32	382	11.5	1511	1 PDR1_YEAST	Q02785 saccharomyc
33	374	11.2	1511	1 PDR5_YEAST	P33302 saccharomyc

ALIGNMENTS

RESULT 1

ABG5_HUMAN STANDARD: PRT; 651 AA.

AC Q9H222;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).

GN ABCG5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID:9606;

RP [1]

RP SEQUENCE FROM N.A., AND VARIANT GLU-604.

RC TISSUE=Liver;

EX MEDLINE-20553648; PubMed-11099417;

RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,

RA Kvitrovich P., Shan B., Barnes R., Hobbs H.H.,

RT "Accumulation of dietary cholesterol in sitosterolemia caused by

RT mutations in adjacent ABC transporters.";

RL Science 290:1771-1775(2000).

RN [2]

RN SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA H-389; H-419 AND P-419,

RP AND VARIANT E-604.

RC TISSUE=Liver;

EX MEDLINE-20578753; PubMed-11138003;

RA Lee M.-H., Lu K., Hazard S., Shulenin S., Hidaka H., Kojima H.,

RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,

RT "Identification of a gene, ABCG5, important in the regulation of

RT dietary cholesterol absorption.";

RL Nat. Genet. 27:79-83(2001).

RN [3]

RN MEDLINE-21474438; PubMed-11590207;

RA Schmitz G., Langmann T., Heimerl S.,

RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";

RL J. Lipid Res. 42:1513-1520(2001).

RN [4]

RN VARIANTS SITOSTEROLEMIA Q-146; H-389; P-419; H-419 AND S-550, AND

RP VARIANT E-604.

EX MEDLINE-21344600; PubMed-11525359;

RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,

RA Ose U., Stalenhoef A.F.H., Meitinen T., Bjorkhem I., Bruckert E.,

RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,

RT "Two genes that map to the STSL locus cause sitosterolemia: genomic

RT structure and spectrum of mutations involving sterolin-1 and

RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";

RL Am. J. Hum. Genet. 69:278-290(2001).

CC -1- FUNCTION: Transporter that appears to play an indispensable role

CC of the enterocytes and in the selective sterol excretion by the

CC liver into bile.

CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to

ABCG8 along a pathway regulating dietary-sterol absorption and excretion.

-1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

-1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels in the small intestine and colon.

-1- DISASE: Defects in ABCG5 are a cause of sitosterolemia, also known as phytosterolemia or shellfish sterolemia, a rare autosomal recessive disorder characterized by increased intestinal absorption of all sterols including cholesterol, plant and shellfish sterols, and decreased biliary excretion of dietary sterols into bile. Sitosterolemia patients have hypercholesterolemia, very high levels of plant sterols in the plasma, and frequently develop tendon and tuberosity xanthomas, accelerated atherosclerosis and premature coronary artery disease.

-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG5 (WHITE) SUPERFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF320293; AAG40003.1; -
EMBL: AF312715; AAG53099.1; -
Genew; HGNC:13886; ABCG5.
MIM: 605459; -
DR MIM: 210250; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR Prodom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism; Disease mutation.

FT DOMAIN 1 383 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 384 404 1 (POTENTIAL).
FT DOMAIN 405 421 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 422 442 2 (POTENTIAL).
FT DOMAIN 443 462 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 463 483 3 (POTENTIAL).
FT DOMAIN 484 503 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 504 524 4 (POTENTIAL).
FT DOMAIN 525 528 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 529 549 5 (POTENTIAL).
FT DOMAIN 550 623 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 624 644 6 (POTENTIAL).
FT DOMAIN 645 651 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 86 93 ATP (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 146 146 E -> Q (IN SITOSTEROLEMIA).
FT VARIANT 389 389 R -> H (IN SITOSTEROLEMIA).
FT FTID-Var.012244.
FT VARIANT 419 419 R -> H (IN SITOSTEROLEMIA).
FT FTID-Var.012245.
FT VARIANT 419 419 R -> H (IN SITOSTEROLEMIA).
FT FTID-Var.012246.
FT VARIANT 419 419 R -> P (IN SITOSTEROLEMIA).
FT FTID-Var.012247.
FT VARIANT 550 550 R -> S (IN SITOSTEROLEMIA).
FT FTID-Var.012248.
FT VARIANT 604 604 O -> E.
FT FTID-Var.012249.
FT SEQUENCE 651 AA; 72503 MM; 950BAFBCBBA1536 CRC64;
FT FTID-Var.012248.
FT FTID-Var.012249.

Query Match 100.0%; Score 3326; DB 1; Length 651;
Best Local Similarity 100.0%; Pred. NO. 1.3e-224;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MGDLSTLPGGSGMGLQVNRGSSSLGAPAPAPAPPHSGILIHASYSHVRPMDITSC 60

DB 1 MGDLSTLPGGSGMGLQVNRGSSSLGAPAPAPAPPHSGILIHASYSHVRPMDITSC 60
61 ROOFTROLIKDVLVYVESQOICITGSSGCKRTLLDAMSGRLAGTFLGENTYNGRL 120
61 ROOFTROLIKDVLVYVESQOICITGSSGCKRTLLDAMSGRLAGTFLGENTYNGRL 120
121 REPOFODCFSYVLSQDTLLSLTRETLYTLLAIIRGNPGSFQKVEAVAEISLSHV 180
121 REPOFODCFSYVLSQDTLLSLTRETLYTLLAIIRGNPGSFQKVEAVAEISLSHV 180
181 ADRLIGNSLIGISNGERRRVSIAOQLQDPVLMFDEPTGLDCMTANOIVLLVEAR 240
181 ADRLIGNSLIGISNGERRRVSIAOQLQDPVLMFDEPTGLDCMTANOIVLLVEAR 240
241 RNRIVVLIHOPRSELFOFDKIALISFGELFCGTPAEMLEDFNDGCPPEHNPDE 300
241 RNRIVVLIHOPRSELFOFDKIALISFGELFCGTPAEMLEDFNDGCPPEHNPDE 300
301 YMDLTSVDPQSKEREIEFSKRVQMESAYKKSACHTLKNIERKHUKTLPMVPEKTRD 360
301 YMDLTSVDPQSKEREIEFSKRVQMESAYKKSACHTLKNIERKHUKTLPMVPEKTRD 360
361 SPGVSKTGLVLRRTNRLVNRKLAIVITRLLQNLIMGFLFVLRVNSVLKGIQDRV 420
361 SPGVSKTGLVLRRTNRLVNRKLAIVITRLLQNLIMGFLFVLRVNSVLKGIQDRV 420
421 GLTYQVATPYTGMLNANVNFPPVLRVNSVDSQDGLQKQOMLALVLAHLPSSVATM 480
421 GLTYQVATPYTGMLNANVNFPPVLRVNSVDSQDGLQKQOMLALVLAHLPSSVATM 480
481 IFSSVCYWTGLHPEVARNFGYFSAALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
481 IFSSVCYWTGLHPEVARNFGYFSAALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
541 GIVVSGFLRNIOENPIFKIISFTPKYCSSEILVNEFGLNFTGSSVSTTNMC 600
541 GIVVSGFLRNIOENPIFKIISFTPKYCSSEILVNEFGLNFTGSSVSTTNMC 600
601 AFTGIGIPIETPCGATSRFTMNFLLTSFPAIVLGIIVFKIRDHLSR 651
601 AFTGIGIPIETPCGATSRFTMNFLLTSFPAIVLGIIVFKIRDHLSR 651

RESULT 2
ABG5_MOUSE STANDARD; PRT; 652 AA.
ID ABG5_MOUSE
AC 099PE8:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=20578753; Pubmed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Alikmets R., Sakuma N., Pegoraro R., Stivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.*
RL Nat. Genet. 27:79-83(2001).
RN [2]
RP TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=20553648; Pubmed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kvitrovich P., Shan B., Barnes R., Hobbs H.H.;
RT Accumulation of dietary cholesterol in sitosterolemia caused by

RT mutations in adjacent ABC transporters.*;
 RL Science 290:1771-1775(2000).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver x receptor/retinoid x receptor (LXR/RXR) pathway.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AF312713; AAG53097.1; -
 CC MGD; MGI:1351659; Abcg5.
 CC InterPro: IPR003593; AAA_Atpase.
 CC Pfam: PF00005; ABC_tran; ABC_transporter.
 CC ProDom: PD000006; ABC_transportr; 1.
 CC SMART; SM00382; AAA; 1.
 CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
 CC ATP-binding; Glycoprotein; Transmembrane; Transport.
 CC DOMAIN 1 385
 CC TRANSMEM 386 406
 CC DOMAIN 407 422
 CC TRANSMEM 423 443
 CC DOMAIN 444 463
 CC TRANSMEM 464 484
 CC DOMAIN 485 504
 CC TRANSMEM 505 525
 CC DOMAIN 526 529
 CC TRANSMEM 530 550
 CC DOMAIN 551 622
 CC TRANSMEM 623 643
 CC DOMAIN 644 652
 CC NP_BIND 67 94
 CC CARBOHYD 410 410
 CC CARBOHYD 585 585
 CC CARBOHYD 592 592
 CC SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;
 SO
 Query Match 82.3%; Score 2738.5; DB 1; Length 652;
 Best Local Similarity 80.1%; Pred. No. 1.3e-183;
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;
 QY 1 MGD:SLTPGSGMGLQVNRGSSGLEAPAPAPPE-HSLGILHASYSVSHRVPMDITS 59
 DB 1 MGE:PLSPGAPPHIRGSLSSLEDSGVTEGRHSLGVLHVSYSNRYGPMWNIS 60
 QY 60 CROQWROIILKDVSLYVESGOIMCIGSSGKTTILDMASGRGRACFTFGEVYVNGRA 119
 DB 61 CQAKNDRIILKDVSLYVESGOIMCIGSSGKTTILDMASGRGRACFTFGEVYVNGRA 120
 QY 120 LRRDQFDCFSYVLOSDFLLSLVRETLTALAIIRGNPGSFQKVEAVMELSLSH 179
 DB 121 LRRDQFDCFSYVLOSDFLLSLVRETLTALAIIRGNPGSFQKVEAVMELSLSH 180
 QY 180 VADRILGNYSILAGISGTERRRVSTAAQLADPKVLFDEPTTGLDCKMTANOIVVLLA 239
 DB 181 VADMIGSYNFGGIGSSGERRRVSTAAQLADPKVLMDEPTTGLDCKMTANOIVVLLA 240

QY 240 RRRRIYVLIHOPRSELFOPLKIALNSGELIFCSTPAEMDFNDGCPPEHNSPND 299
 DB 241 RRRRIYVLIHOPRSELFOPLKIALNSGELIFCSTPAEMDFNDGCPPEHNSPND 300
 QY 300 FYWDLTSVDTOSKERIEFTSKRQVOMESAKYKSAJCHTKLNIERKHKLTLPMPPEFK 359
 DB 301 FYWDLTSVDTOSKERIEFTSKRQVOMESAKYKSAJCHTKLNIERKHKLTLPMPPEFK 360
 QY 360 DSEPGSKGLVLRVTRNLVNRKLAVITRLLQNLIMGLFLLFVLRVNSVNLKGAIDOR 419
 DB 361 DSEPGSKGLVLRVTRNLVNRKLAVITRLLQNLIMGLFLLFVLRVNSVNLKGAIDOR 420
 QY 420 VGLTLVQVATPYTGLNVLNVPVLRVNSVNSDSDGLYQKQNMALAVLPPSVAT 479
 DB 421 VGLTLVQVATPYTGLNVLNVPVLRVNSVNSDSDGLYQKQNMALAVLPPSVAT 480
 QY 480 MIFSSVCYWTGLLHPEVARGFYSALALPHLIGEFLLVLYGIYONPINSVALLSI 539
 DB 481 VIFSSVCYWTGLLHPEVARGFYSALALPHLIGEFLLVLYGIYONPINSVALLSI 540
 QY 540 AGVLVSGFLRNIOEPIFKIISYFFQKYSGLVNEFYGLNFTGSSNVSVTNPM 599
 DB 541 SGILIGSGFIRNIOEPIFKIISYFFQKYSGLVNEFYGLNFTGSSNVSVTNPM 600
 QY 600 CAPTQIOPIETKTCGATSRFTMNFLLXSFPAVYIGIVFKIRDLISR 651
 DB 601 CAITQVOPITEKTCGATSRFTMNFLLXSFPAVYIGIVFKIRDLISR 652
 RESULT 3
 ID ABG5_RAT STANDARD; PRT; 652 AA.
 AC 099PE7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
 GN ABCG5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Small Intestine;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulentin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RT Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption.*;
 RL Nat. Genet. 27:79-83(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AF312714; AAG53098.2; -
 CC InterPro: IPR003593; AAA_Atpase.

DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran.1;
 DR ProDom: P000006; ABC_transportr.1;
 DR SMART: SM00382; AAA.1;
 DR ProSITE: PS00211; ABC_TRANSPORTER.1;
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 386 406 1 (POTENTIAL).
 FT TRANSMEM 407 422 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 423 443 2 (POTENTIAL).
 FT TRANSMEM 444 463 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 464 484 3 (POTENTIAL).
 FT TRANSMEM 485 504 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 505 525 4 (POTENTIAL).
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 530 550 5 (POTENTIAL).
 FT TRANSMEM 551 624 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 625 645 6 (POTENTIAL).
 FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
 FT NR BIND 87 94 ATP (POTENTIAL).
 FT CAROHRD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHRD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 652 AA; 73342 MW; 4D42FE2BAB0AD59 CRC64;

Query Match 82.0%; Score 2727.5; DB 1; Length 652;
 Best Local Similarity 79.4%; Pred. No. 7.7e-183;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

1 MGDLSTLPGGSMKQVNRGSSQSSLEGAPATAPAP-PSGLILHASYSHRYPMWMDTS 59
 1 MGELPFLSPGARPHNRGSSQSSLEGAPATAPAP-PSGLILHASYSHRYPMWMDTS 60
 60 CROOITROILKDVSLVYSGQIMCILLSSGSGKTTLDAMSGLRAGTFLGEVYVNGRA 119
 61 COOKMDKRLDVSLVYSGQIMCILLSSGSGKTTLDAMSGLRAGTFLGEVYVNGRA 120
 120 LRROFOCCFYSYVOSDTLSSLVYRETLTALTAIRGPGSFQKVEAVMAELSLH 179
 121 LRROFOCCFYSYVOSDTLSSLVYRETLTALTAIRGPGSFQKVEAVMAELSLH 180
 180 VADRILGNSLGGISYGERRRVSIADQLDQPKVMEDEPTTGDCMTANOIVLVELA 239
 181 VADOMIGNYNGGSSIGERRRVSIAADQLDQPKVMEDEPTTGDCMTANOIVLVELA 240
 240 RRRNIVVITIQPSESEFQFLDKALISFGLICGCPAEKMLDPFNDGCGPCHESNPF 299
 241 RRRNIVVITIQPSESEFQFLDKALISFGLICGCPAEKMLDPFNDGCGPCHESNPF 300
 300 FYMDLTSVDQSKREIEFTSKRVOMISGAKSAICHTLKNIRMKMLKTLPMVPEKTK 359
 301 FYMDLTSVDQSKREIEFTSKRVOMISGAKSAICHTLKNIRMKMLKTLPMVPEKTK 360
 360 DSPGVFSLGVLRRVTRNLVTRNLVTRNLVTRNLVTRNLVTRNLVTRNLVTRNLV 419
 361 NPPCMFCKLGVLLRRVTRNLVTRNLVTRNLVTRNLVTRNLVTRNLVTRNLVTRNL 420
 420 VGLLYQFVATPYGCHLNAVNLFPVLAHSVQESQDGLYQKQWMLALALVLPSSVAT 479
 421 VGLLYQFVATPYGCHLNAVNLFPVLAHSVQESQDGLYQKQWMLALALVLPSSVAT 480
 480 MIFSSVCYMTLGLPEVARFGYFSALLAPHLIGEFITLVGLVQNPNTVNSVALIST 539
 481 VIFSSVCYMTLGLPEVARFGYFSALLAPHLIGEFITLVGLVQNPNTVNSVALIST 540
 540 AGVLVSGFLNIDEMPIPKIISYTFQKYSCEILVNEFYGLNFGSSNVSVTPNM 599
 541 SGLIGSGFINIEMPIPKIISYTFQKYSCEILVNEFYGLNFGSSNVSVTPNM 600
 600 CAPTQGIQTEKTCPGATSRPTNMFLLISYTPALVILGIVKTRDHLISR 651
 601 CSMTOGIQTEKTCPGATSRPTNMFLLISYTPALVILGIVKTRDHLISR 652

RESULT 4
 ID ABG8_MOUSE STANDARD: PRT; 673 AA.
 AC Q9DBM0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (SterolIn-2).
 GN ABCG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Silvastava A.K.,
 RA Patel S.B.;
 RA "Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolIn-1 and
 sterolIn-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai I., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl L.M., Lewis S., Matsuo Y., Mikado I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staib L.F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberees P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilmink L.,
 RA Wyshav-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [3]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kvtirovich P., Shan B., Barnes R., Hobbs H.H.;
 RA "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (by similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoid X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.

CC -1- CAUTION: Seems to have a defective ATP-binding region.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF324495; AAK84079.1; -
 DR EMBL: AK004871; EMB23630.1; -
 DR MGI: 1914720; Abcg8.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transporter.1.
 DR ProSITE: PS00211; ABC_TRANSPORTER.1.
 KM Glycoprotein; Transmembrane; Transport; Alternative splicing.
 FT DOMAIN 1 413
 FT TRANSMEM 414 434
 FT DOMAIN 435 447
 FT TRANSMEM 448 468
 FT DOMAIN 469 496
 FT TRANSMEM 497 517
 FT DOMAIN 518 526
 FT TRANSMEM 527 547
 FT DOMAIN 548 569
 FT TRANSMEM 570 590
 FT DOMAIN 591 639
 FT TRANSMEM 640 660
 FT DOMAIN 661 673
 FT CAROHD 619 619
 FT VARSPLIC 377 377
 SQ SEQUENCE 673 AA: 75995 MW: 7801261.5 Df589 CRC64;
 Query Match 21.0%; Score 698; DB 1; Length 673;
 Best Local Similarity 28.7%; Pred. No. 3.2e-41;
 Matches 194; Conservative 133; Mismatches 264; Indels 84; Gaps 18;
 Oy 11 GSMGVNRGSSQSSLEGAPAPAPPHSGILHASTSVSHVR-----PMD-I 57
 Db 14 GTVLQDASQGLQDSL-----FSESDNSLYFTYSGQSNLEVRDLTYQVDIASQVPMFQRL 69
 Oy 58 TSCROOQTRQI-----LKDVSLYVESGOIMCITLSSGSGKTLTLDAMSGRLGRATF 109
 Db 70 AQFKPRMSSHSQSCSEIGIRNLSFKVNSGOMLIGSSGGRASLDVITGR-GHGKM 128
 Oy 110 -LGEVYNGRRLRRQPODCFSYVLSQPTLLSLTVRETLHYTALLAI-RGNGPQSFQK 167
 Db 129 KSGQWINGQPTQLVRYKCAVAVRHQDLPLNLTRETLAFIQMRPLRFESQARKR 188
 Oy 168 VEAVMAELSLHVDRLIGNSLGISTGERRRVSIAQLQDPKVMLEDPPTGLDCMT 227
 Db 189 VEDVIAELRLQCANTRVNTYRGVSGGERRRVSIGQLMNGIILIDEPGSLDFT 248
 Oy 228 ANQIVLVLLVLRRIYVLTIRPSELFPQLEKIALISGELIFCSTPRMELDFDC 287
 Db 249 AHNVTTLRLAKGNRLVLSLHOPRSDIFLFLVLTMTSGTPIYLGAAQAOVQYFTSI 308
 Oy 288 GYPCPEHSNPDPFYMDLTSVDQSKREIETSKRVOMIESAYKKA-----ICHKTLKN 341
 Db 309 GHPCPRYSNPADFYVDLTSIDRSKREVAVERKAQSLALFLKQVGFDFLKAKEKE 368
 Oy 342 IERKHKLTLPVVFTRKDS-----PGVSKLGLVLRRTNVLVKNKLAIVTRLLQN 393
 Db 369 LMTSTHVSITLT-----TQDTCGTAVELPGMIEQSTLIRQISNDRDPTLLIHSEA 424
 Oy 394 LIMGLFLEFLYLRVSNVTKGAIQDRVGLLYQFQCAPPYGMANVULFYLRAVSDQES 453
 Db 425 CIASLIIGFLYTGGAQOL--SPMDTALLFMGALIPFNVTLDVYKCHSERSMLYVEL 482
 Oy 454 QDGLYQWQWMLALVALHVLFPFSVATMIFSSVCWTGLHPFVARPFYFSAILAPHLIG 513

Db 483 EQLTAGYFFFAKILGELPEHCAYIITAMPIYLTNRPPVEL-----LL--HELL 534
 Oy 514 EFL-----TVLLIGIVONPNI-VNSVALLSIAGLVGSGFLRNIOEMPIPKIISYTF 566
 Db 535 VMLVVECCRTMALAASAMLPFTFMSSEFCNALYNSFYLAGFMINLDNIMIVAWTSKLS 594
 Oy 567 FQKYSCEILVVEFGLNFT--CGSSNVSTTPKCAFTQGIQFIKTCRGATSRFTNMF 624
 Db 595 FLKWCSSGLMOIQFONCHLVTYTOIGNFTFSLIGDTM-----ISANDLMS 637
 Oy 625 LILYSFIPALVILGI 639
 Db 638 HPLXALY--LIVIGI 650
 RESULT 5
 ABG8_HUMAN STANDARD; PRT; 673 AA.
 ID ABG8_HUMAN 09H21:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
 GN ABG8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RP [1]
 RP SEQUENCE FROM N.A., VARIANTS STIGSTEROLEMIA T-231; Q-263; R-574 AND
 RP R-596, AND VARIANT C-54.
 RP MEDLINE=2053648; PubMed=11099417;
 RA Berge R.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwaterovich P., Shan B., Barnes R., Hobbs H.H.;
 RA "Accumulation of dietary cholesterol in sitosterolemia caused by
 RA mutations in adjacent ABC transporters.";
 RA Science 290:1771-1775(2000).
 RL [2]
 RL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS STIGSTEROLEMIA H-184;
 RL T-231; Q-263; H-405; P-501; S-543; P-572; E-574; R-574; R-596 AND
 RL F-570 DEL, AND VARIANTS H-19; C-54; K-238; V-259; K-400; R-575 AND
 RL A-632.
 RC TISSUE=Liver;
 RC MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RA "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RA structure and spectrum of mutations involving sterolin-1 and
 RA sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RA Am. J. Hum. Genet. 69:278-290(2001).
 RL [3]
 RL REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Helmerl S.;
 RA "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RA J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are
 CC produced by alternative splicing. Isoform 2 is a minor form
 CC detected in approximately 10% of the cDNA clones.
 CC -1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon. Detectable in a wide variety of
 CC human tissues.
 CC -1- DISEASE: Defects in ABG8 are a cause of sitosterolemia, also

RESULT 6
ABCG8_RAT
ID ABCG8_RAT STANDARD: PRT: 672 AA.
AC P58428;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Metlinen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA Patel S.B.;
RT *Two genes that map to the SRS1 locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterolin-1 and
RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.*;
RT Am. J. Hum. Genet. 69:278-290(2001).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -1- CAUTION: Seems to have a defective ATP-binding region.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF351785; AAK84831.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR pfam: PF00005; ABC_transporter.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT DOMAIN 1 412
FT TRANSMEM 413 433
FT DOMAIN 434 446
FT TRANSMEM 447 467
FT DOMAIN 468 495
FT TRANSMEM 496 516
FT DOMAIN 517 525
FT TRANSMEM 526 546
FT TRANSMEM 547 568
FT DOMAIN 569 589
FT TRANSMEM 590 629
FT TRANSMEM 629 649
FT DOMAIN 650 672
FT TRANSMEM 672 692
FT CARBOHYD 618 618
FT VARSPLIT 376
FT SEQUENCE 672 AA: 75878 MW: 60889735F3FA4C36F CRC64;
Query Match 20.8%; Score 691; DB 1; Length 672;
Best Local Similarity 28.3%; Pred. No. 9.8e-41;

Matches 169; Conservative 126; Mismatches 264; Indels 88; Gaps 15;
QY 23 SSLEGAPANAPEPHSLGILHASYSVSHRYR-----PW-----MDITSC 60
DB 21 SSLQDSVFSESDNSLYFTYSGQSNLEVDLYDYQDMAQVPMFQALQFKLPMHSRGS 80
QY 61 RQMTQOILKDVSLYVESQIMKILSSGSGKTTLLDANSGRIGRGFTLGEVYVNGRAL 120
DB 81 QDSMDLGI-RNLSFKVSGQMLAIGSAGGRATLLDYTGRRHGKMRSGQIWMGOPS 139
QY 121 RREQFQDCFSYVQSTPLSLVRETLHTALALRGNPGSF-----QKKEVMAEL 175
DB 140 TQQLDLCYAHVAGQDQLPMLNLTVELTFTIQRNL-----PKTESQAQRKRDVEDYIAEL 195
QY 176 SLSHVADRILIGNSLIGISTGERRRYSIAAQLDQPKVMLEDEPTTGDCMTANQIVLL 235
DB 196 RLQOCANTRYGNTYVGVSGGERRRYSIGVQLMNPGLILDEPTSGDSFTAHNLVRL 255
QY 236 VELARRNRYVLTIRHPRSELPOLPKIALISFGEILFCGTPAEMLDENDCGPPERS 295
DB 256 SRLAKGNRLVLTSLHPRSDIFRLFDVLVLTGTPYIGVAOHMVOYFTSIGYPCPRYS 315
QY 296 NPFDFYMDLTSVDTSKEREIETSKRWOMIESAYKKS-----ICKTLKNIERKKHLK 349
DB 316 NPADFVDTSLIDRSKEDQVATMERARLIALFLERKVGDFDEFLKAEKSLD-----TG 371
QY 350 TLPMPPEKTKDS-----PGVFSKIGVLRVRYTNLVNKLAVITRLQLNGLFTLL 401
DB 372 TYAVSQTLNQDNFCGTAELPGLIQFTLLIRQINDRDPLTLTHGAEKMLSLIIG 431
QY 402 FVYLKRSVNLKGAIDRGLQYFGAFTPYGMALNVLFPYLAVSQDSQDGLYQRM 461
DB 432 FLYYGHADRL--SFMDMAALLFMIGALIPFNVLIDVVKCHSERSLYLEEDGLYTG 489
QY 462 QMVLVYALHLPFSVYATMIFSSCYWTGLHPEVARPFYSALLAPLIGEL----- 516
DB 490 PYEFKAVIGELDEHCAVYLLIGMPYITWLTNLR-----GPELHFMMLMLVFPCC 541
QY 517 -TLVLGLIVQNPNV-VNSVALLSLAGVVGSGFLRNIOEMPIPEKIIISYTFQKYSCEI 574
DB 542 RTMALAASAMLPFTFHMSFCCNALNVSFTLTFAGFMNLNMLIVPMWIKMFLRNCFSG 601
QY 575 LVVNEFYGLNFCGSSNVSVTNPMCAFTQGIQTEKTCPG--ASRFTMNLILYSFTP 632
DB 602 LMQIOFNHITYTTOIGNLTFV-----PGDAMVTAMDLSNPILYAIY- 643
QY 633 ALVIIGI 639
DB 644 -LIVIGI 649
RESULT 7
ABG2_HUMAN
ID ABG2_HUMAN STANDARD: PRT: 655 AA.
AC Q9DNU0; Q95374; Q9NDS0; Q9BR73;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-
DE binding cassette transporter) (Breast cancer resistance protein).
GN ABCG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99065313; PubMed=9850061;
RA Alikmets R., Schmitt L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCG) on
Cancer Res. 58:5337-5339(1998)."

RN [12] SEQUENCE FROM N.A.
 RP TISSUE-Breast cancer;
 RC MEDLINE-99080071; PubMed-9861027;
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RT "A multidrug resistance transporter from human MCF-7 breast cancer
 cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
 RN [13] ERRATUM.
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
 RN [14] SEQUENCE FROM N.A.
 RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,
 RA Sugimoto Y.;
 RT "Breast cancer resistance protein constitutes a 140-kDa complex as a
 homodimer";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [15] SEQUENCE OF 198-655 FROM N.A.
 RP TISSUE-Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Makamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Niinomiya K., Iwayanagi T.;
 RT "NDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [16] REVIEW.
 RP MEDLINE-21474438; PubMed-11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
 IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
 CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME
 RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN. DISPLAY
 DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
 MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF103796; AAD09188.1;
 DR EMBL: AF098951; AAC97367.1;
 DR EMBL: AB056867; BAB39212.1;
 DR EMBL: AK002040; BAA92050.1;
 DR Genev: HGNC:74; ABCG2.
 DR KIM: 603756;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transpfr.
 DR Pfam: PF00005; ABC_tran.1.
 DR Prodom: PD000006; ABC_transpfr.1.
 DR SMART: SM00382; AAA.1
 DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE NEG.
 KW ATP-binding; transmembrane; transport.
 FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 396 416 POTENTIAL.
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 429 449 POTENTIAL.
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 478 498 POTENTIAL.
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 507 527 POTENTIAL.
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 536 556 POTENTIAL.
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 80 87 ATP (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
 FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
 FT CONFLICT 208 208 F -> S (IN REF. 1).
 FT CONFLICT 315 316 MISSING (IN REF. 5).
 FT CONFLICT 482 482 R -> T (IN REF. 2).
 SQ SEQUENCE 655 AA; 72343 MM; 89A6D351IDC5CE0 CRC64;

Query Match 20.3%; Score 676.5; DB 1; Length 655;
 Best Local Similarity 29.0%; Pred. No. 9.8e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SSSLECAPATAP---EPHSIGILHASYSVSHRPPMDITSCROQWTRILKQSYLYE 77
 DB 13 SQGNTNGEPRAVSDNKAFTGAVLSPNITCYRKLSGFLPCKRPVEKELTSLNGIMK 72
 QY 78 SGQIMCLIGSSGSKRTLLDMSGRGLAGATFLGEVYNGRNLRRPODFSVYASDT 137
 DB 73 PG-LNALIGPYGCKSSLDVLAARKDPSG-LSGDVLING-APPANPKNSGYVODV 129
 QY 138 LLSLTVRETLHTALLAIRGNPG-SQKQVEAVMAELSHVADRLIGYSLAGISTG 196
 DB 130 VMGLTVRENLQFSAAALRLATMTNHEKNERINIVIELGLDKVADSKVGTQPIRGVSG 189
 QY 197 ERRRVSTAAQLLOPKYVALFDEPTTGLDCAANOIVYLVLAERNRRIYVLTIHOPSEL 256
 DB 190 ERKRTSTIGMELTPSLIFLDEPTTGLDSTANNVLLKMKSKQGRITLFSHPRISI 249
 QY 257 FQLPDKTALLSFGELIFCGPAEMLDEFNDGYPCEPHSNPFYMDLTSVDTO-----SK 312
 DB 250 FKLPDSLTLLASGLMFGHPAQELAGFEASGYCEAVNNPADFLDINDSTVAVALNR 309
 QY 313 ERE-----LETSR-----VOMISAKKSKALCKT-----LNIERMKHLKTLPMVPF 356
 DB 310 EEDFKATEILPESKQDKPLIEKLAETIYVNSSEFYETRAEDHQLSGEKKKITYFKETISY 369
 QY 357 KTKDSPGVFSKGLVLRVTRNLRNKLAVTRLLQMLINGLFL--LEFVLRVRSVNLKG 414
 DB 370 TT-----SFHQQLRNVRSKRSKNLGNQASIAQIIVYVGLVIGALIFGLKNDST----- 421
 QY 415 AIQDRVGLLYQFGATPYTGLNLANVNLFPYLRAVSDQSDGLYQKQMMALVAL-HYLP 473
 DB 422 GIONRAVLFL-FLTTNOCFSSVSVAVELFVVEKKLFIEHYISGYRVSSYFLGRLSLDLP 480
 QY 474 FSVYATMIFSSVCWTGLGHPVARFGYFSAALLAPHLIEFLTVLLGIQONINYSV 533
 DB 481 MRMLPSITIFCYIVFMGLPKADAFVMMFTLL--MVAYSASMAALAAAGOSVSV 537
 QY 534 VALSLAGV--LVQSGFLRNIOEMPIFKIISYTFQKYSCEILIVNEFVGLNCGSSN 591
 DB 538 TLMTICFVMMIFSGILVNLITIASLSLQVLFSTIRYGFALQHNHEFLGQNFPCG--- 594
 QY 592 VSVYTNPMCAFTGQIOFIKTCG 615
 DB 595 LNAIGNPNCAVY-----TCYG 610
 RESULT 8
 YOH5_YEAST STANDARD; PRT; 1294 AA.

AC 008234; 008233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97321807; PubMed-9178509;
 RX Tzeremia M., Katsoulou C., Alexandraki D.;
 RT Sequence analysis of a 33.2 kb segment from the left arm of yeast
 RT chromosome XV reveals eight known genes and ten new open reading
 RT frames including homologues of ABC transporters, inositol
 RT phosphatases and human expressed sequence tags.*;
 RL Yeast 13:383-389(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 DR EMBL: 274817; CAA99085.1;
 DR EMBL: 274816; CAA99084.1;
 DR SGD: S0005435; YOL075C.
 DR InterPro: IPR003593; AAA_Arpase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 KM
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 605 625 POTENTIAL.
 FT TRANSMEM 1039 1059 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1267 1287 POTENTIAL.
 FT NP_BIND 62 69 ATP (POTENTIAL).
 FT NP_BIND 727 734 ATP (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 983 983 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1294 AA; 145157 MW; C555500A45E9284E CRC64;

Query Match 18.9%; Score 627; DB 1; Length 1294;
 Best Local Similarity 31.7%; Pred. No. 6,7e-36;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

OY 65 TROLIKDVSLVESGOICLSSGSGKTTLLDANGSLGRAGTFL-----GEVYVNRRA 119
 DB 706 TKELIYVNAIFKRGIMVIMPGSSGSSLLNLISGL KSSFAKEDTSGSIFNDIQ 764
 OY 120 LRROPODCSYVLSQST-LISLTVRETLHYTALLAIRGNPGSFQKVEAVNAELSLIS 178
 DB 765 VSELIFKNCVSYQDDHLIALTLVETLKYAALRLAHILTEARMERNTNLRISLOLK 824

OY 179 HVADRILGNVSLGISTGERRRYSIAQALLODPKWLFDPEPTGLDCMTANQVLVLLVEL 238
 DB 825 HCENNITIGNEFVNGISGGERRRYTMGOVLNDPILLDEPTSGLSFSTATILELEKL 884
 OY 239 AR-RNRIVVLTITIQPSELFLQFLDKIAIIS-FCGLFCGPPAMLPFNDCCGRCDEHSN 296
 DB 885 CROGQTIITITHPSELFLKRGVNLAKSRTPANSPDEMIATFELGNCSEFTN 944
 OY 297 PEDFYMDLTSVDTSKRELETSKRVQMESAYKSAICHTLKINTERKMLTLPMPVP 356
 DB 945 VADFFDLISVTQNQNEISRAVEKILISAK-----ANNDN-ESLSPTISER 994
 OY 357 KTRDSGVSKGLVLLRRTNRY-----RNLAVITRLQNLINLGLFL 401
 DB 995 QOYSQSFTEYSEFYRK-PANDVLAVYVNVKROFTTTRSPDSLMARIAQLPGLGVIRA 1053
 OY 402 FEVLVRSNVLKGAIQDRVGLLYQFVGATPYGMNAVMLFPLRAVSOESQDGLYOKM 461
 DB 1054 LEFAPVKHNT--SISNRLGLQEST-ALIEVGMGLNLCYPERDYEEYENDVYGLA 1110
 OY 462 OMMLAYALHLPESVAVATMIFSSVCYTGLLHPEVARFEGFSALLAPHLI--GEFLTL 518
 DB 1111 PEFLLAVMTLEPLSLASVLAIVFVYLAGL-PRTA--GNFATVYCSPIVYCCGRRLG 1167
 OY 519 VLLGIYQNNN-IVNSVVALSLAGLVSGFLRNIOEMPPEKIIISYFTFOKYSILVY 577
 DB 1168 MNTFEFERGFVANCISIIISIGTQMSGMLST-----GMSRVLKGNYLNPVGYTSMIIN 1223
 OY 578 NEFYG-LNFTC--GSSNVSVTNPMTCAFTQG 605
 DB 1224 FAPPGMLKLTCEGDKNSDGT-----CEPANG 1250

RESULT 9
 WHIT_LUCCU
 ID WHIT_LUCCU STANDARD; PRT; 677 AA.
 AC 005360;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
 OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
 OC Muscomorpha; Oestroidea; Calliphoridae; Lucilia.
 OX NCBI_TaxID=7375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97087158; PubMed-8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.;
 RT "The structure, sequence and developmental pattern of expression of
 RT the white gene in the blowfly *Lucilia cuprina*.";
 RL Insect Mol. Biol. 5:251-260(1996).
 RN [2]
 RP SEQUENCE OF 490-584 FROM N.A.
 RA MEDLINE-90264941; PubMed-1971656;
 RA Elizur A., Vacek A.T., Howells A.J.;
 RT "Cloning and characterization of the white and topaz eye color genes
 RT from the sheep blowfly *Lucilia cuprina*.";
 RL J. Mol. Evol. 30:347-358(1990).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

OY 388 TRLLONLIMGFLFEVLKNSVNLKALIDRVGL-----LYOFVGAFTPYGMNAYNL 441
 DB 442 VALLQOTAMVATLI-----GSIYFGVILDDQDGMVNINGSLFLFLTNFTQNVAVINV 493
 OY 442 FVVLKAVSDQESODLXOKWOMLAVLHVLPFSVAVAMIFSSVCYTWLGLHPEARREY 501
 DB 494 FSAELPVLKREKRSRIYRDYFELCKTAEPLFAVPEVFSIYPMGL-----RTG- 547
 OY 502 FSAALLAPHLIGEFTLVLLGIQVQPN-----INVSVALLSIA-----GYVCSG 547
 DB 548 -----ATHYL-----TTLFTVIVLVANVSTSFGLLSCASSISMSALVCPVYIPLIFGG 598
 OY 548 FLRNIOEMPIPKIISYTFYFQYCEIIVNNEFYGL-----NFTCGSSNVSVTT 596
 DB 599 FFLNLSAVPAYFKYLSYLSMFRYANEDALLINQMSVYDQELACTANVTCPREIILET 657

RESULT 12
 WHIT_DROME STANDARD: PRT: 687 AA.
 ID WHIT_DROME STANDARD: PRT: 687 AA.
 AC P10090: Q9V3A2; Q9YX33;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE White protein.
 GN W OR EG: BACN33B1.1 OR CG2759.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Head;
 RX MEDLINE=90221897; PubMed=2109311;
 RA Repling M., Mount S.M.;
 RT "Sequence of a cDNA from the Drosophila melanogaster white gene.";
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
 RT "DNA sequence of the white locus of Drosophila melanogaster.";
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Miwa S.,
 RA Yamamoto D.;
 RT "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blasej R.G., Zhang C., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandaril D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris M.,

RA Harris M.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisli D., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Mobarry C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paolel J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Split E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.R., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Dreano S., Gloux S., Leilaure V., Mottier S., Galibert F., Borkova D.,
 RA Minena B., Katalos F.C., Louis C., Siden-Klamos I., de Pablos B.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Wenner M., Moutflet F.,
 RA Belnert N., Dove G., Schaefer U., Jackle H., Bucheton A.,
 RA Callister D.M., Campbell I.A., Parlanitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valent P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 RN [6]
 RP SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=89339145; PubMed=2503416;
 RA Tearle R.G., Belote J.M., McKown M., Baker B.S., Howells A.J.;
 RT "Cloning and characterization of the scarlet gene of Drosophila
 RT melanogaster.";
 RL Genetics 122:595-606(1989).
 CC -!- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY
 CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
 CC RESPONSIBLE FOR EYE COLOR. WHITE DIVERGIZE WITH BROWN FOR THE
 CC TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF
 CC TRYPTOPHAN.
 CC -!- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X51749; CAA36038.1; -
 CC EMBL: X02974; CAA26716.1; -
 CC EMBL: AB028139; BAA78210.1; -
 CC EMBL: AE003425; AAF45826.1; -
 CC EMBL: AL133506; CAB65847.1; -
 CC EMBL: X76202; CAA53795.1; -
 CC PIR: S07263; EYFEM;
 CC PIR: S08635; SYFEM;
 CC FlyBase: FBgn0003996; w.
 CC InterPro: IPR003593; AAA_ATPase.
 CC InterPro: IPR003439; ABC_transporter.

```

DR InterPro: IPR005284; Pigment-permease.
DR Pfam: PF00005; ABC_tran; 2.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfam: TIGR00955; 3a01204; 1.
DR Prosite: PS00211; ABC_TRANSPORTER; 1.
DR Pigment: ATP-binding; Transmembrane; Transport.
KW NP_BIND 130 137 ATP (BY SIMILARITY).
FT TRANSMEM 435 453 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
FT TRANSMEM 515 533 POTENTIAL.
FT TRANSMEM 542 563 POTENTIAL.
FT TRANSMEM 576 594 POTENTIAL.
FT TRANSMEM 659 678 POTENTIAL.
FT CONFLICT 25 29 GDSGA -> LIFEIPHYCHVTAD (IN REF. 2 AND 3).
FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).
FT CONFLICT 335 371 VQAQCPNTNPADPFYQVLAAYVAGREIESRRIARIC ->
FT ITLHNSYPAAWPSVLPPTTIRRTFTYRCWPLCPDGRSSSVI
FT GSPRYG (IN REF. 3).
SQ SEQUENCE 687 AA: 75672 MW: 24AFAD799DE0D396 CRC64:

Query Match 18.1%; Score 602.5; DB 1; Length 687;
Best Local Similarity 28.8%; Pred. No. 1.5e-34;
Matches 180; Conservative 131; Mismatches 220; Indels 95; Gaps 19;

OY 66 ROLLDVSLYVESGOIMCIISSGSGKTTLDMSGR--LGNAGFLGEEVYNGRARE 123
DB 110 KHLKNCGVAFPGELAVMGSSGKTTLLNALFRSPQGVQSPSGRLNGPVDK 169
OY 124 QPQDFSYVLOSDTLSSLYVAETLHYTLAIRGNPSGSPK--KYEAVAEELSHV 180
DB 170 EQMAGCAVYQDDLFISLTAREHLIFQAM--VRRPHLYTRQARAVDQVIOELSLKC 227
OY 181 ADRLLG-KYSLGISTGERRRVSTIAQLQDPKVALFDEPTTGLDCMTANQIVLLVELA 239
DB 228 QHTIIGVGRKGLSGGERKRLAFSEALTDPLICEDEPTSGDSEFAHSAVQVIAKLKS 287
OY 240 RNRRLVLTIHOPRESELPOLFKAILSEFELFCGTPAEMDFPDGCGPCPEHSNPD 299
DB 288 QKQKVIITLIHOPSELEFELDKILMAEGRAFLGTFSEAVDFSYGAQCPNTYNPAD 347
OY 300 FYMDITSVDQSKEREIEFTSKRVOMIESAYKKSALCHTKLNIEKMLKTLPMVPFTK 359
DB 348 FYVQVLAV---VPGREIESRRIARICDNPAS---KVANDMEQLATKNE---KPL 396
OY 360 DSP-----GVSKLGVLLRRYTRNLVRNKLAVITLLQNLINGFLV-LEFVLVRSN 410
DB 397 EDPNGYTYKATWQFRAVLRMSLVAKLEPLVYKRLIQTTWVAIIIGLIFLGQQLTQ 456
OY 411 VLKGAIDORVGLTQVGFATPYTGMLNVLNLPVLRASDQESODGLQKQOMLAVLH 470
DB 457 V---GVANINAIPLFLNMTFQNFATINFTSELPFMEARSLKRCOTYFGKTIA 513
OY 471 VLPSVAVATMFSSVCYWTGLGHEVARRGYFSAALLAPHLIGELFVLIGIVONPIV 530
DB 514 ELPLFLVPLVETALAYPMIGLVLF-----FNCIALVTIV--ANVS 556
OY 531 NSVALLSIAG-----VLVSGELRNIOENPIPKIITYTFQKYCEI 574
DB 557 TSFGILLISCASSSTSMALSVGPVYIIPFLIFGFGFLNSGVPVLYKLMSYLSWFRYANEG 616
OY 575 LVNVEFYGL---NFTGSSNVSVTNPMCAFTQGIQIEKCP--GASRRTMNFLLYS 629
DB 617 LLINQMAVDEGEISCTSSNT-----TQSSGKVIETLNSA--A 655
OY 630 FIP---ALVTLGIVFKIRDLISR 651
DB 656 DLPIDYGLAIL-IVSFRVLAYLALR 680

RESULT 13
ABGI_HUMAN

```

DR	EMBL	AF3289151	CAC00730.1	JOINED
DR	EMBL	AF323658	AAK28836.1	-
DR	EMBL	AF323644	AAK28836.1	JOINED
DR	EMBL	AF323645	AAK28836.1	JOINED
DR	EMBL	AF323646	AAK28836.1	JOINED
DR	EMBL	AF323647	AAK28836.1	JOINED
DR	EMBL	AF323648	AAK28836.1	JOINED
DR	EMBL	AF323649	AAK28836.1	JOINED
DR	EMBL	AF323650	AAK28836.1	JOINED
DR	EMBL	AF323651	AAK28836.1	JOINED
DR	EMBL	AF323652	AAK28836.1	JOINED
DR	EMBL	AF323653	AAK28836.1	JOINED
DR	EMBL	AF323654	AAK28836.1	JOINED
DR	EMBL	AF323655	AAK28836.1	JOINED
DR	EMBL	AF323656	AAK28836.1	JOINED
DR	EMBL	AF323657	AAK28836.1	JOINED
DR	EMBL	AF323664	AAK28842.1	-
DR	EMBL	AF323658	AAK28833.1	-
DR	EMBL	AF323640	AAK28833.1	JOINED
DR	EMBL	AF323645	AAK28833.1	JOINED
DR	EMBL	AF323646	AAK28833.1	JOINED
DR	EMBL	AF323647	AAK28833.1	JOINED
DR	EMBL	AF323648	AAK28833.1	JOINED
DR	EMBL	AF323649	AAK28833.1	JOINED
DR	EMBL	AF323650	AAK28833.1	JOINED
DR	EMBL	AF323651	AAK28833.1	JOINED
DR	EMBL	AF323652	AAK28833.1	JOINED
DR	EMBL	AF323653	AAK28833.1	JOINED
DR	EMBL	AF323654	AAK28833.1	JOINED
DR	EMBL	AF323655	AAK28833.1	JOINED
DR	EMBL	AF323656	AAK28833.1	JOINED
DR	EMBL	AF323657	AAK28833.1	JOINED
DR	EMBL	AF323660	AAK28838.1	JOINED
DR	EMBL	AF323663	AAK28841.1	ALT_INIT
DR	EMBL	AF323658	AAK28835.1	JOINED
DR	EMBL	AF323642	AAK28835.1	JOINED
DR	EMBL	AF323645	AAK28835.1	JOINED
DR	EMBL	AF323646	AAK28835.1	JOINED
DR	EMBL	AF323647	AAK28835.1	JOINED
DR	EMBL	AF323648	AAK28835.1	JOINED
DR	EMBL	AF323649	AAK28835.1	JOINED
DR	EMBL	AF323650	AAK28835.1	JOINED
DR	EMBL	AF323651	AAK28835.1	JOINED
DR	EMBL	AF323652	AAK28835.1	JOINED
DR	EMBL	AF323653	AAK28835.1	JOINED
DR	EMBL	AF323654	AAK28835.1	JOINED
DR	EMBL	AF323655	AAK28835.1	JOINED
DR	EMBL	AF323656	AAK28835.1	JOINED
DR	EMBL	AF323657	AAK28835.1	JOINED
DR	EMBL	AF323662	AAK28840.1	-
DR	EMBL	AF323643	AAK28837.1	JOINED
DR	EMBL	AF323645	AAK28837.1	JOINED
DR	EMBL	AF323646	AAK28837.1	JOINED

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21518231; PubMed-11606068;
RA Engel T., Lorkowski S., Lueken A., Rust S., Schluter B., Berger G.,
RA Cullen P., Asseman G.,
RT "The human ABCG4 gene is regulated by oxytocin and retinoids in
RT monocyte-derived macrophages."
RL Biochem. Biophys. Res. Commun. 288:483-488(2001).
RN [2]
RP SEQUENCE OF 20-646 FROM N.A.
RC Tissue-dorsal root ganglion;
RA Oldfield S., Lowry C.A., Lightman S.L.;
RT "Cloning and expression of a novel mammalian white family
RT ABC-transporter: WHITE2."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in macrophage lipid homeostasis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AJ300465; CAC17140.1; -.
DR Genbank: HGNC:13884; ABCG4.
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran: 2.
DR ProDom: PD000006; ABC_transport: 1.
DR SMART: SM00382; AAA: 1.
DR PROSITE: PS00211; ABC_TRANSPORTER: 1.
DR ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN 1 393
FT TRANSMEM 394 414
FT DOMAIN 415 425
FT TRANSMEM 426 446
FT DOMAIN 447 472
FT TRANSMEM 473 493
FT DOMAIN 494 503
FT TRANSMEM 504 524
FT DOMAIN 525 532
FT TRANSMEM 533 553
FT DOMAIN 554 617
FT TRANSMEM 618 638
FT DOMAIN 639 646
FT NP_BIND 102 109
FT CARBOHYD 422 422
SQ SEQUENCE 646 AA: 71895 MW: 90CECE6150772611 CRC64;
Query Match 17.4%; Score 578.5; DB 1; Length 646;
Best Local Similarity 27.1%; Pred. No. 6.6e-33;
Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;
OY 33 PEPSSGLHASYSVSHRVRWMDITSCROQMTQIILKDVSLYVSGQIMCIGSSGSGK 92
DB 54 PKRSADVLEFVELSYVAEGPCW-----RRKGYKTLKLGKFCRRRELIGIMPSGAGK 108
OY 93 TTLDAASGRLAGTGEVYVNGRALRRQFODCFYSYVQSDTLLSLVRETLHTA 152
DB 109 STEFNILAGY--RESGKGGILLVNGRPRELTFKMSCTIQQDMLPLHLYLRAAMVSA 166
OY 153 LLAIRGNPGSGFQKKEAV-----MAELSLSHVADRIGNYSLGISTGERRRVSIAAQ 206
DB 167 NLKLSERQ---EVKKELVETILTALGIMSCSHRTAL-----LSGGQKRLATALE 214
OY 207 LLDPPKVMLEDEPTGLDCMTANQIVLVLELARRNRIVLTTHQPSSELFQLPDKTAL 266

DB 215 LVNPPVAFDEPTGLDASACFQVVSILKSLAOGRTIICTIHQPSAKLFEMPDKLYL 274
OY 267 SEGELIFGCTPAEMDEFNDGPGCPHNSNPFDDYMDLSTVDTSKKEIETSKRVOMIE 326
DB 275 SOGQCTFQVNTNLLPYLKGJLRCPYHNADPDLIEVASG-----EYGLNPLPLF 325
OY 327 SAYKSAICHKHTLKNIEPRKHLKTLPMVPFRTKSP-----GVFSKGLVLLRRVTRN 378
DB 326 RAV-QNGICAAAEKSSPEKNEVAPCPCPPEVDPIESHFASTLTQPCILKRTFLS 384
OY 379 LVNRKLAIVTRLLNGLFLFVLRVRSNVLKGAIDRVGLYQVAGATPYTGMLNA 438
DB 385 ILRDVTLHLRFMSHVIVGLIGLYLHIGDASK--VENNTGCLFESMLFLMRALMPT 442
OY 439 VNLPPVLAAYSDQSDLYOKMOMLALVLPFSVATMIFSSCYTWTGLHPEYAR 498
DB 443 VLTPPLEAVMREHLNTWYSLKAYLLAKTADVDPEQVVCVYCSITYWNTGQPAETSR 502
OY 499 FGYSALLAHLIGEFLTVLIVGIVONPNIVNSVVALSLAGVYNGSGFLRNTQEMP 558
DB 503 FLIFSALATATALVAQSLGL-LIGASNSLOVATFVGVTALPVLLSGFVSFKTYPT 561
OY 559 FKIIISYFPQYCEILVYNEFGD--NFTCGSSNVSVTNPMCAFTOGIQTLEKCPG 615
DB 562 LQMSYLSYVRYGEGVILT-IYGEREDLTC-----LEKCPREP-QSILRALDV 611
OY 616 ATRPTMFLFLYSITPALTGLIVYERKIR 645
DB 612 EDARKYMDFLVIGIFLALRLALVLAFLYR 641

Search completed: June 11, 2003, 09:12:37
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:04:17 ; Search time 41 Seconds
(Without alignments)
2115.760 Million cell updates/sec

Title: US-09-989-981a-6

Sequence: 1 MCDLSSLRTRGSGMGLVNRG.....PALVIIGIVFKIRDLISR 651

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3336	100%	651	AAU96984	Human ABCG5 protei
2	3326	100%	651	AAE13290	Human sitosterolae
3	3323	99.9	651	AAU96992	Human ABCG5 mutant
4	3321	99.8	651	AAU96989	Human ABCG5 mutant
5	3319	99.8	651	AAU96990	Human ABCG5 mutant
6	3319	99.8	651	AAU96993	Human ABCG5 mutant
7	2748.5	82.5	652	AAE13309	Human ABCG5 protei
8	2744.5	82.5	652	AAE13289	Human sitosterolae
9	2742.5	82.5	652	AAE13308	Human sitosterolae
10	2738.5	82.5	652	AAU96985	Mouse ABCG5 protei

11	2727.5	82.5	652	AAU96986	Rat ABCG5 protei
12	2081	62.6	408	AAU96991	Human ABCG5 protei
13	1387.5	41.7	340	AAU96987	Human ABCG5 protei
14	1234	37.1	243	AAU96988	Human ABCG5 protei
15	722	21.7	144	AAU96989	Human ABCG5 protei
16	682.5	20.5	655	AAU96990	Human ABCG5 protei
17	682.5	20.5	655	AAU96991	Human ABCG5 protei
18	680.5	20.5	655	AAU96992	Human ABCG5 protei
19	680.5	20.5	655	AAU96993	Human ABCG5 protei
20	680.5	20.5	655	AAU96994	Human ABCG5 protei
21	680.5	20.5	655	AAU96995	Human ABCG5 protei
22	676.5	20.3	655	AAU96996	Human ABCG5 protei
23	674.5	20.3	655	AAU96997	Human ABCG5 protei
24	674.5	20.3	655	AAU96998	Human ABCG5 protei
25	674.5	20.3	655	AAU96999	Human ABCG5 protei
26	672.5	20.2	655	AAU97000	Human ABCG5 protei
27	660	19.8	657	AAU97001	Human ABCG5 protei
28	602.5	18.1	602	AAU97002	Human ABCG5 protei
29	600	18.0	602	AAU97003	Human ABCG5 protei
30	588.5	17.7	648	AAU97004	Human ABCG5 protei
31	587.5	17.7	635	AAU97005	Human ABCG5 protei
32	587.5	17.7	632	AAU97006	Human ABCG5 protei
33	581.5	17.5	687	AAU97007	Human ABCG5 protei
34	578.5	17.4	646	AAU97008	Human ABCG5 protei
35	578	17.4	604	AAU97009	Human ABCG5 protei
36	578	17.4	604	AAU97010	Human ABCG5 protei
37	575.5	17.3	649	AAU97011	Human ABCG5 protei
38	575.5	17.3	653	AAU97012	Human ABCG5 protei
39	571	17.2	609	AAU97013	Human ABCG5 protei
40	569.5	17.1	666	AAU97014	Human ABCG5 protei
41	564.5	17.0	616	AAU97015	Human ABCG5 protei
42	553	16.6	666	AAU97016	Human ABCG5 protei
43	515.5	15.5	832	AAU97017	Human ABCG5 protei
44	473	14.2	689	AAU97018	Human ABCG5 protei
45	463.5	13.9	812	AAU97019	Human ABCG5 protei

ALIGNMENTS

RESULT 1	AAU96984	standard; Protein; 651 AA.
ID	AAU96984	
AC	AAU96984	
DT	30-JUL-2002	(first entry)
DE	Human ABCG5 protei	
XX		
KW	Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; chromosome 2p21.	
KW	Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; chromosome 2p21.	
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 2..15	/note="Encoded by GRCYC"
XX		
PD	04-APR-2002.	
XX		
PF	25-SEP-2001; 2001MO-US29859.	
XX		
PR	25-SEP-2000; 2000US-235268P.	
XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA	(PATE/) PATEL S B.	
PA	(DEAN/) DEAN M.	
XX		
PI	Patel SB, Dean M.	

XX WPI; 2002-416483/44.
 DR N-PSDB; ABR51681.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 XX sitosterolemia, arteriosclerosis and heart diseases
 PS Claim 52; Page 35-36; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX
 XX
 SQ Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSSLPFGSGMGLQVNRGSSSLGAPATAPAPHSGLTHASVSHRRVPMDDTISC 60
 DB 1 MGDLSSLPFGSGMGLQVNRGSSSLGAPATAPAPHSGLTHASVSHRRVPMDDTISC 60
 QY 61 RQOWTROLKLVSLYVESGQIMCITGSSGKTTLLDMSGRGLGAGTFLEVVYNGRAL 120
 DB 61 RQOWTROLKLVSLYVESGQIMCITGSSGKTTLLDMSGRGLGAGTFLEVVYNGRAL 120
 QY 61 RQOWTROLKLVSLYVESGQIMCITGSSGKTTLLDMSGRGLGAGTFLEVVYNGRAL 120
 DB 61 RQOWTROLKLVSLYVESGQIMCITGSSGKTTLLDMSGRGLGAGTFLEVVYNGRAL 120
 QY 121 RREOFODCFYVLOSDTLLSLVRETHTYALLAIRGNPGSPQKKEAVMAELTSHV 180
 DB 121 RREOFODCFYVLOSDTLLSLVRETHTYALLAIRGNPGSPQKKEAVMAELTSHV 180
 QY 121 RREOFODCFYVLOSDTLLSLVRETHTYALLAIRGNPGSPQKKEAVMAELTSHV 180
 DB 121 RREOFODCFYVLOSDTLLSLVRETHTYALLAIRGNPGSPQKKEAVMAELTSHV 180
 QY 181 ADRLIGNYSLGISTGERRRVSIQAQLQDPKVMLEDEPTGLDQMTANQIVLLVELAR 240
 DB 181 ADRLIGNYSLGISTGERRRVSIQAQLQDPKVMLEDEPTGLDQMTANQIVLLVELAR 240
 QY 241 RNRIVVLTIRPRLSTLPOLFKIALISGELIFGCTPAEMLDFFNDCGYPCEHNSNPDF 300
 DB 241 RNRIVVLTIRPRLSTLPOLFKIALISGELIFGCTPAEMLDFFNDCGYPCEHNSNPDF 300
 QY 241 RNRIVVLTIRPRLSTLPOLFKIALISGELIFGCTPAEMLDFFNDCGYPCEHNSNPDF 300
 DB 241 RNRIVVLTIRPRLSTLPOLFKIALISGELIFGCTPAEMLDFFNDCGYPCEHNSNPDF 300
 QY 301 YMDLTSVDTOSKEREIETSKRVMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 DB 301 YMDLTSVDTOSKEREIETSKRVMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 QY 301 YMDLTSVDTOSKEREIETSKRVMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 DB 301 YMDLTSVDTOSKEREIETSKRVMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 QY 361 SPGVFSKILGVLRLRVTRNLVNRKLAIVTRLQNLIMGLFLEFVLRAVSANLKAIDRV 420
 DB 361 SPGVFSKILGVLRLRVTRNLVNRKLAIVTRLQNLIMGLFLEFVLRAVSANLKAIDRV 420
 QY 421 GLTQFGATPYTGMNAVNLFPVLRAVSDQSDGLYQKQMMALAYALHVLPSVATM 480
 DB 421 GLTQFGATPYTGMNAVNLFPVLRAVSDQSDGLYQKQMMALAYALHVLPSVATM 480
 QY 481 IFSSVCWTTLGLHPEVNAFGEFSAALLAPHLIGEFITLVLTGYONRNIVSVALLSIA 540
 DB 481 IFSSVCWTTLGLHPEVNAFGEFSAALLAPHLIGEFITLVLTGYONRNIVSVALLSIA 540

QY 541 GVLVSGFRLNIOEMPPEKTIISYTFPKYCSLIVWVEFGNLNCCSSNVSYTTPMC 600
 DB 541 GVLVSGFRLNIOEMPPEKTIISYTFPKYCSLIVWVEFGNLNCCSSNVSYTTPMC 600
 QY 601 AFTGCIQPIETKPCGATSRFTWNLILVSFIPALVILIVFKIRIDHLISR 651
 DB 601 AFTGCIQPIETKPCGATSRFTWNLILVSFIPALVILIVFKIRIDHLISR 651

RESULT 2
 AAE13290
 ID AAE13290 standard; Protein; 651 AA.
 AC AAE13290;
 DT 12-FEB-2002 (first entry)

Human sitosterolemia susceptibility gene (SSG) protein.

Human; sitosterolemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.

OS Homo sapiens.

W0200179272-A2.
 25-OCT-2001.

18-APR-2001; 2001WO-US12758.

18-APR-2000; 2000US-198465P.

15-MAY-2000; 2000US-204234P.

(TULIA-) TULARIK INC.

Tian H, Schultz J, Shan B;

WPI; 2002-017598/02.

N-PSDB; AAD22009.

Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of Ssg polypeptide for treating
 PT sterol-related disorder

Claim 19; Fig 8; 105pp; English.

The invention relates to an isolated sitosterolemia Susceptibility Gene
 (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG protein. Human SSG is located on chromosome
 CC 2p21.

Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGDLSSLPGGSMGLQVNRGSSSLLEGAPATAPAPPHSGILHASTSVSHRVPMDITSC 60
DB 1 MGDLSSLPGGSMGLQVNRGSSSLLEGAPATAPAPPHSGILHASTSVSHRVPMDITSC 60
OY 61 RQWRTQILKDVSLVYSGOIMCIIIGSSGSKTTLLDMSGRGLRAGTFLGEVYNGRAL 120
DB 61 RQWRTQILKDVSLVYSGOIMCIIIGSSGSKTTLLDMSGRGLRAGTFLGEVYNGRAL 120
OY 121 RREPODFCFYVLSDDLSSLTVEETLHTATLAIIRGNPSSPKKYEAVAEISLSHV 180
DB 121 RREPODFCFYVLSDDLSSLTVEETLHTATLAIIRGNPSSPKKYEAVAEISLSHV 180
OY 181 ADRLIGNSLGIGISTGERRRYSIAAQLDPPKMLFDEPTTGIDCMANQIVLVLELAR 240
DB 181 ADRLIGNSLGIGISTGERRRYSIAAQLDPPKMLFDEPTTGIDCMANQIVLVLELAR 240
OY 241 RNRIVVLTIHQPRSELQPLDKIALISFGELIFCGTPEAMLDFFNDGCPCEHSNPDF 300
DB 241 RNRIVVLTIHQPRSELQPLDKIALISFGELIFCGTPEAMLDFFNDGCPCEHSNPDF 300
OY 301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHTKLNTERKHLKTLPMVPFKTKD 360
DB 301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHTKLNTERKHLKTLPMVPFKTKD 360
OY 361 SPGVSKLGVLLRRVTNMLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420
DB 361 SPGVSKLGVLLRRVTNMLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420
OY 421 GLTQFGATPYTGMLNANVLPVLRVAVSDSDGIXOKOMLAVALAVLPESVATM 480
DB 421 GLTQFGATPYTGMLNANVLPVLRVAVSDSDGIXOKOMLAVALAVLPESVATM 480
OY 481 IFSSVCTWTGLHPEVNAFGYSAAALAPHLIGELTLVLLGIYONPNIIVSVALLSIA 540
DB 481 IFSSVCTWTGLHPEVNAFGYSAAALAPHLIGELTLVLLGIYONPNIIVSVALLSIA 540
OY 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSELIIVNEFGLNFTGSSNVSATYTNMC 600
DB 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSELIIVNEFGLNFTGSSNVSATYTNMC 600
OY 601 AFTOGIOPIEKTCPGATSRFTMNFLLIYSFIPALVILGIIVFKIRDLHISR 651
DB 601 AFTOGIOPIEKTCPGATSRFTMNFLLIYSFIPALVILGIIVFKIRDLHISR 651

```

RESULT 3
AA096992
ID AA096992 standard; Protein: 651 AA.
AC AA096992;
XX
DT 30-JUN-2002 (first entry)
DE Human ABCG5 mutant E146Q protein sequence.
XX
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW mutant; muteln.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key location/Qualifiers
FT Misc-difference 146 /note- "wild-type Glu substituted by Gln".
XX
PD 04-APR-2002.
PF 25-SEP-2001; 2001WO-0529859.
XX

```

PR 25-SEP-2001; 2000US-235268P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
PI Patel SB, Dean M;
XX WPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
PS Claim 12; Page -- 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant E146Q protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
CC of the specification.
XX
XX Sequence 651 AA:
XX
XX Query Match 99.9%; Score 3323; DB 23; Length 651;
XX Best Local Similarity 99.8%; Pred. No. 0;
XX Matches 650; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

1 MGDLSSLPGGSMGLQVNRGSSSLLEGAPATAPAPPHSGILHASTSVSHRVPMDITSC 60
1 MGDLSSLPGGSMGLQVNRGSSSLLEGAPATAPAPPHSGILHASTSVSHRVPMDITSC 60
61 RQWRTQILKDVSLVYSGOIMCIIIGSSGSKTTLLDMSGRGLRAGTFLGEVYNGRAL 120
61 RQWRTQILKDVSLVYSGOIMCIIIGSSGSKTTLLDMSGRGLRAGTFLGEVYNGRAL 120
121 RREPODFCFYVLSDDLSSLTVEETLHTATLAIIRGNPSSPKKYEAVAEISLSHV 180
121 RREPODFCFYVLSDDLSSLTVEETLHTATLAIIRGNPSSPKKYEAVAEISLSHV 180
181 ADRLIGNSLGIGISTGERRRYSIAAQLDPPKMLFDEPTTGIDCMANQIVLVLELAR 240
181 ADRLIGNSLGIGISTGERRRYSIAAQLDPPKMLFDEPTTGIDCMANQIVLVLELAR 240
241 RNRIVVLTIHQPRSELQPLDKIALISFGELIFCGTPEAMLDFFNDGCPCEHSNPDF 300
241 RNRIVVLTIHQPRSELQPLDKIALISFGELIFCGTPEAMLDFFNDGCPCEHSNPDF 300
301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHTKLNTERKHLKTLPMVPFKTKD 360
301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHTKLNTERKHLKTLPMVPFKTKD 360
361 SPGVSKLGVLLRRVTNMLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420

DE Human ABCG5 mutant R389H protein sequence.

XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutlein.

XX Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 389 /note- "Wild-type Arg substituted by His"
 FT FT
 XX WO200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 XX 25-SEP-2000; 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;
 PI PI
 XX WPI; 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PT PT
 XX Claim 7; Page -: 66pp; English.

XX The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R389H protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 CC XX
 SQ Sequence 651 AA;

Query Match 99.8%; Score 3321; DB 23; Length 651;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGDLSTPGSGMGLVNRGSSSLGAPATAPPHSLGLTHASYSVSHRVPMWDTSC 60
 DB 1 MGDLSTPGSGMGLVNRGSSSLGAPATAPPHSLGLTHASYSVSHRVPMWDTSC 60
 QY 61 ROGWTRQILKDVSLVYSSQIMCITLSSSGSKTTLTLDAMSGRLGRAGTFAGYVNGRAL 120
 DB 61 ROGWTRQILKDVSLVYSSQIMCITLSSSGSKTTLTLDAMSGRLGRAGTFAGYVNGRAL 120

QY 121 RREPOFQDESVYLOSDTLSSITVRETIHTYATALLAIRGNPGSPQKYEAVMAELSLSHV 180
 DB 121 RREPOFQDESVYLOSDTLSSITVRETIHTYATALLAIRGNPGSPQKYEAVMAELSLSHV 180
 QY 181 ADRLIGNSYLGIISTGERRRVSIQAQLQDPKVMLEDEPTTGLODMTANQIVLVLELAR 240
 DB 181 ADRLIGNSYLGIISTGERRRVSIQAQLQDPKVMLEDEPTTGLODMTANQIVLVLELAR 240
 QY 241 RNRIVYLTIHQPRSELFDLFDKIALISGELIFCGTPAEMLDFNDCGYPCEHSNPDEF 300
 DB 241 RNRIVYLTIHQPRSELFDLFDKIALISGELIFCGTPAEMLDFNDCGYPCEHSNPDEF 300
 QY 301 YMDLTSVDTOSKERIEFTSKRVOMESAYKKSALCHTKLNIEKKHLKTLPMVPFKTKD 360
 DB 301 YMDLTSVDTOSKERIEFTSKRVOMESAYKKSALCHTKLNIEKKHLKTLPMVPFKTKD 360
 QY 361 SPGVFSKLGVLRLRYTRMLVNRKLAIVIRLQNLIMGLFLFVLRVSNVLRKALIDRV 420
 DB 361 SPGVFSKLGVLRLRYTRMLVNRKLAIVIRLQNLIMGLFLFVLRVSNVLRKALIDRV 420
 QY 421 GLTIQFVATPYTGMLNANVLFVLRVAVDSQDGLYQKQMMALAVLAPPSVATM 480
 DB 421 GLTIQFVATPYTGMLNANVLFVLRVAVDSQDGLYQKQMMALAVLAPPSVATM 480
 QY 481 IFSSVCYWTGLHPEVNAFGEFSAAALLAPHLIGFELTVLGIYQNPINVSVALLSIA 540
 DB 481 IFSSVCYWTGLHPEVNAFGEFSAAALLAPHLIGFELTVLGIYQNPINVSVALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVNVFEGNLFTGSSNVSTYTNMC 600
 DB 541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVNVFEGNLFTGSSNVSTYTNMC 600
 QY 601 AFTGIGQFIETKCPGATSRFTMNLILYSFIPALVILGIYVEKIRDLHLSR 651
 DB 601 AFTGIGQFIETKCPGATSRFTMNLILYSFIPALVILGIYVEKIRDLHLSR 651

RESULT 6
 ID AA096993 standard; Protein; 651 AA.
 AA096993;
 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419P protein sequence.
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutlein.
 XX Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 419 /note- "Wild-type Arg substituted by Pro"
 FT FT
 XX WO200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 XX 25-SEP-2000; 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;

DR WPI: 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide; and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT xanthoma, arteriosclerosis and heart diseases
 XX
 PS Claim 10; Page -: 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419P protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX
 XX Sequence 651 AA:
 Query Match 99.8%; Score 3319; DB 23; Length 651;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MODLSLTPGSGMGLQVNRGSSQSLGAPATAPEPHSLGIIILASYSVSRVAPMDITSC 60
 DB 1 MODLSLTPGSGMGLQVNRGSSQSLGAPATAPEPHSLGIIILASYSVSRVAPMDITSC 60
 QY 61 RQOMTROLIKDVSILVYESGOIMCIISSGSGKTTLLDMNSGRLAGTFLGEVYNGRAL 120
 DB 61 RQOMTROLIKDVSILVYESGOIMCIISSGSGKTTLLDMNSGRLAGTFLGEVYNGRAL 120
 QY 121 RREOPDCEVYVLOSDTLLSLTVEETLHTYLLAIRGNPSSFOKKEVAVAEISLSHY 180
 DB 121 RREOPDCEVYVLOSDTLLSLTVEETLHTYLLAIRGNPSSFOKKEVAVAEISLSHY 180
 QY 181 ADRLIGNYSIGISTGERRRVSAIAQLLODPKVMLEDEPTTLLDCTANQIYVLLVELAR 240
 DB 181 ADRLIGNYSIGISTGERRRVSAIAQLLODPKVMLEDEPTTLLDCTANQIYVLLVELAR 240
 QY 241 RNRIVVLTTHOPRSELPFLDKITAILSFCELIFCGTPAEMLDPFNDGCGPCHEHNPDE 300
 DB 241 RNRIVVLTTHOPRSELPFLDKITAILSFCELIFCGTPAEMLDPFNDGCGPCHEHNPDE 300
 QY 301 YMDLSVDPQSKEREIETSKRYOMIESAYKKSALCHKTLKNIERKHKLTLPVFPKTKD 360
 DB 301 YMDLSVDPQSKEREIETSKRYOMIESAYKKSALCHKTLKNIERKHKLTLPVFPKTKD 360
 QY 361 SPGVSKLGLVLRVTRNLVNRKLAIVITRLQNLIMGLFLFVLRVSNVYKGIADRV 420
 DB 361 SPGVSKLGLVLRVTRNLVNRKLAIVITRLQNLIMGLFLFVLRVSNVYKGIADRV 420
 QY 421 GLIYFVGATPTGMLNANVLPVLRVAVSDQSGLYOKQWOMLAVLHVLFPFSVATM 480
 DB 421 GLIYFVGATPTGMLNANVLPVLRVAVSDQSGLYOKQWOMLAVLHVLFPFSVATM 480
 QY 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGELTIVLGIYONPNIVNSVYALLSIA 540
 DB 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGELTIVLGIYONPNIVNSVYALLSIA 540

DB 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGELTIVLGIYONPNIVNSVYALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYFFQKCEILVNVNEFGNFTGSSNSVYTPNPMC 600
 DB 541 GVLVSGFLRNIOEMPIPKIISYFFQKCEILVNVNEFGNFTGSSNSVYTPNPMC 600
 QY 601 AFTGIGIETKTCPGATSRFTMNFILYSFIPALVILGIVKIRIDHLISR 651
 DB 601 AFTGIGIETKTCPGATSRFTMNFILYSFIPALVILGIVKIRIDHLISR 651
 RESULT 7
 ID AAE13309 standard; Protein; 652 AA.
 XX AAE13309;
 AC AAE13309;
 XX
 DT 12-FEB-2002 (first entry)
 DE Mouse sitosterolemia susceptibility gene (SSG) protein variant #2.
 DE
 KW Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis; mutant;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; mutant;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.
 XX
 OS Mus sp.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 28
 FT
 XX
 PN WO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001WO-US12758.
 XX
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 XX
 DR WPI: 2002-017598/02.
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 XX
 PS Disclosure: Page -: 105pp; English.
 CC The invention relates to an isolated sitosterolemia susceptibility gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Gly28 with Ala.
 CC Note: The present sequence is not shown in the specification but is

CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289) and
CC shown in figure 7 of the specification.
CY

SQ Sequence 652 AA;

Query Match	82.68;	Score 2748.5;	DB 23;	Length 652;
Best Local Similarity	80.48;	Pred. No. 1.1e-280;		
Matches 524; Conservative	64;	Mismatches 63;	Indels 1;	Gaps 1;

QY	1	GGDLSLTPGSGMGLQVNRGSSSLSECAPATAPPE-BSLGLIHMAYSVSHRYRPMWIDTS	59
Db	1	MELEPESPEGRGPHINRKSLSLEQASVTCETEARSHSLGYLHVSVSNRYRPMWNIKS	60
QY	60	CROQPTROLKDVSLYVESGOIMCTLSSGSGKTTLLDAMSGRLGRATFGEYVYNGRA	119
Db	61	COQKNDROLKDVSLYTESGOIMCTLSSGSGKTTLLDASGRLLRTTETLGEVEFNCE	120
QY	120	LAREPODCFSYVLOSPTLLSLYRETLHYTALAIIRGNPGSPQKVEAVAMELSH	179
Db	121	LARDPODCFSYVLOSPTLLSLYRETLHYTALAIIRGNPGSPQKVEAVAMELSH	180
QY	180	VADRLIGNSLGGISGTGERRRVSAQAQLQDPKVMLEDEPTGLDCKMANQIVYLLA	239
Db	181	VADQMGISNFGISGGERRRVSAQAQLQDPKVMLEDEPTGLDCKMANQIVYLLA	240
QY	240	RNRIVYLTIHQPSRELQOLFDKIALISFGLIFCGTAEMLDPFNDCCGYCPCHSNPD	299
Db	241	RNRIVYLTIHQPSRELQOLFDKIALISFGLIFCGTAEMLDPFNDCCGYCPCHSNPD	300
QY	300	FYMDLTSVDTSKRESEIETSKRVOMIESAVKKSATCHTLNIEPMKHLKTLPMVPETK	359
Db	301	FYMDLTSVDTSKRESEIETSKRVOMIESAVKKSATCHTLNIEPMKHLKTLPMVPETK	360
QY	360	DSPGVSKLGVLLRRVTRNLVNRKLAVITRLLQNLIMGLFLFYLRYRSNVLKGAIDR	419
Db	361	DSPGVSKLGVLLRRVTRNLVNRKLAVITRLLQNLIMGLFLFYLRYRSNVLKGAIDR	420
QY	420	VGLLYQFPGATPYTGMLNAVNLFEYLRVAVSQESQDGLYQKQMMALAHVLPFSVAT	479
Db	421	VGLLYQFPGATPYTGMLNAVNLFEYLRVAVSQESQDGLYQKQMMALAHVLPFSVAT	480
QY	480	MISSVCWTGLLPEVARFGYFSALLAPHLISEPFLVYLGIVQNDNYNSVALLSI	539
Db	481	VIFSSVCWTGLLPEVARFGYFSALLAPHLISEPFLVYLGIVQNDNYNSVALLSI	540
QY	540	AGVLVSGFLNIOEMPIFKLIISYTFEOKKCESELVNVNFEYGLNFCGSSNVSVTNPM	599
Db	541	SGLLIGSEFINIOEMPIFKLIISYTFEOKKCESELVNVNFEYGLNFCGSSNVSVTNPM	600
QY	600	CAPFQGIQFIKTCPCGATSRFTNMFLLYSFIPALVILGIVFKIRIDHLSR	651
Db	601	CATQGVQFIEKTCPCGATSRFTNMFLLYGRIPALVILGIVFKIRIDHLSR	652

RESULT 8

AAEI3289

AAEI3289 standard; Protein; 652 AA.

AAEI3289;

12-FEB-2002 (first entry)

Mouse stlrosterolaemia susceptibility gene (SSG) protein.

Mouse; stlrosterolaemia susceptibility gene; SSG; atherosclerosis; steroid-related disorder; hyperlipidemia; hypercholesterolaemia; gall stone; coronary heart disease; cardiovascular disease; arthritis; xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

Mus sp.

MO200179272-A2.

PD 25-OCT-2001.

PF 18-APR-2001; 2001WO-US12758.

PR 15-MAY-2000; 2000US-204234P.
PR 18-APR-2000; 2000US-198465P.
PR

PA (TULA-) TULARIK INC.

PI Tian H, Schultz J, Shan B;

DR WPI; 2002-017598/02.
DR N-PSDB; AAD22008.
DR

PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder -

PS Claim 19; Fig 7; 105pp; English.

The invention relates to an isolated sitosterolemia susceptibility gene (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP) binding cassette (ABC) family cholesterol transporter. SSG is useful for identifying a compound useful in the treatment or prevention of a sterol-related disorder, including sitosterolemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or nutritional deficiencies. SSG is also useful for treating cholesterol-associated diseases or conditions including coronary heart disease and other cardiovascular diseases, and sitosterolemia-associated condition including arthritis, xanthomas and chronic haemolytic anaemia. SSG expression cassette is useful in the production of transgenic non-human animals. SSG genes and their homologues are useful as tools for a number of applications including diagnosing sitosterolemia and other cardiovascular disorders, for forensics and paternity determinations, and for treating any of a large number of SSG associated diseases. The present sequence is mouse SSG protein. Mouse SSG is located on chromosome 17.

SQ Sequence 652 AA;

Query Match	82.5%	Score	2744.5	DB	23	Length	652
Best Local Similarity	80.2%	Pred. No.	2.9e-280				
Matches	523	Conservative	64	Mismatches	64	Indels	1
				Gaps			1

[illegible]

QY 420 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGKXOKQMMLAYALHVPESVAT 479
 DB 421 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGKXOKQMMLAYALHVPESVAT 480
 QY 480 MIFSSVCYWTGLGHEVAFRGYFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLGHEVAFRGYFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 540
 QY 540 AGVLVSGFLRNIOEMPIPKIISYFTFOKCYSEILVNEFGNLFTCGSSNVSVTNP 599
 DB 541 SGLLIGSGFIRNIOEMPIPKIISYFTFOKCYSEILVNEFGNLFTCGSSNVSVTNP 600
 QY 600 CAPTGOIPIETKTCGATSRFTMNLILYSPALVILGIYVKIDHLISR 651
 DB 601 CAITGVQFIETKTCGATSRFTMNLILYSPALVILGIYVKIDHLISR 652

RESULT 9
 AAEL3308
 ID AAEL3308 standard; Protein: 652 AA.
 AC AAEL3308;
 DE 12-FEB-2002 (first entry)
 XX Mouse sitosterolemia susceptibility gene (SSG) protein variant #1.
 XX
 KW Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis; mutant;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; mutant;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.
 XX Mus sp.
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 17
 FT /note= "Wild type Ile substituted with Leu"
 PN W02001/9272-A2.
 PD 25-OCT-2001.
 PF 18-APR-2001; 2001WO-US12758.
 XX
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 DR WPI: 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 XX
 PS Disclosure: Page -: 105pp; English.
 XX
 CC The invention relates to an isolated Sitosterolemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number

CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Ile17 with Leu.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAEL3308) and
 CC shown in figure 7 of the specification.

XX Sequence 652 AA:

Query Match 82.5%; Score 2742.5; DB 23; Length 652;
 Best Local Similarity 80.2%; Pred. No. 4.7e-280;
 Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

QY 1 MGDLISLTPGSGAGLVNNGSOSLSGATATAPR-HSIGIILHASYSHRRPMDIIS 59
 DB 1 MGELPFLSPGARGPHLNRSGLSLDGGSVGTGTEARSHVLSVSYSGRPPMWNIS 60
 QY 60 CROQWTRQILKDVSLYVESQIMCIIGSSGSKTTLIDMAGSLGAGTFLGEVYVNGR 119
 DB 61 COQKMDROLKDVSLYIESGQIMCIIGSSGSKTTLIDMAGSLGAGTFLGEVYVNGE 120
 QY 120 LRREPQDCFSYVLOSSTLSSLYVRETLHYTALLAIRNGPSFOKKEAVYAAELISLH 179
 DB 121 LRDDQFQDCFSYVLOSDFLSSLYVRETLHYTALLAIRNGPSFOKKEAVYAAELISLH 180
 QY 180 VADRLIGNSLGIGSGERRRVSIAQILQDDKVMFLDEPTGLOMTNOIVLLIYELA 239
 DB 181 VADMIGSNFGIGSGERRRVSIAQILQDDKVMFLDEPTGLOMTNOIVLLIYELA 240
 QY 240 RRRNRIVLTIHOPRSELPQDFPKIATLSPEGLIFCQTPAEMLDFNDGCPPEHSNP 299
 DB 241 RRRRIYIYTIHQPRSLFQHPFKIATLITGELFCQTPPEMGLFNNCGYPCPEHSNP 300
 QY 300 FYMDLTSVDQSKEREIETSKRVQIMESAYKSAICHKTLKNIERKHILKTLPPVPEKTK 359
 DB 301 FYMDLTSVDQSKEREIETSKRVQIMESAYKSAICHKTLKNIERKHILKTLPPVPEKTK 360
 QY 360 DSPGVSKGLVLRRTNRLVNRKLAIVTRLLONIMGLFLFPLVRAVSNTLKGAVDOR 419
 DB 361 DPPGVSKGLVLRRTNRLVNRKLAIVTRLLONIMGLFLFPLVRAVSNTLKGAVDOR 420
 QY 420 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGKXOKQMMLAYALHVPESVAT 479
 DB 421 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGKXOKQMMLAYALHVPESVAT 480
 QY 480 MIFSSVCYWTGLGHEVAFRGYFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLGHEVAFRGYFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 540
 QY 540 AGVLVSGFLRNIOEMPIPKIISYFTFOKCYSEILVNEFGNLFTCGSSNVSVTNP 599
 DB 541 SGLLIGSGFIRNIOEMPIPKIISYFTFOKCYSEILVNEFGNLFTCGSSNVSVTNP 600
 QY 600 CAPTGOIPIETKTCGATSRFTMNLILYSPALVILGIYVKIDHLISR 651
 DB 601 CAITGVQFIETKTCGATSRFTMNLILYSPALVILGIYVKIDHLISR 652

RESULT 10
 AA096985
 ID AA096985 standard; Protein: 652 AA.
 AC AA096985;
 DE 30-JUL-2002 (first entry)
 XX
 XX Mouse ABCG5 protein.
 DE
 KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX

OS	Mus sp	Location/Qualifiers
XX	Key	Misc-difference 638..652
XX	FT	/note="Encoded by CTag"
XX	XX	W0200227016-A2.
XX	PD	04-APR-2002.
XX	XX	25-SEP-2001; 2001WO-US29859.
XX	XX	25-SEP-2000; 2000US-235268P.
XX	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	PA	(PATE/) PATEL S B.
XX	PI	(DEAN/) DEAN M.
XX	PI	Patel SB, Dean M;
XX	DR	WPI: 2002-416483/44.
XX	XX	N-PSDB; ABR51684.
XX	XX	Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
XX	XX	nucleic acid encoding the polypeptide, useful for treating
XX	XX	sicklesterolemia, arteriosclerosis and heart diseases
XX	XX	Example 3; Page 42; 66pp; English.
XX	XX	The present invention relates to a new mammalian ATP-binding cassette
XX	XX	gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
XX	XX	predisposition for developing sicklesterolemia, arteriosclerosis or heart
XX	XX	disease. The molecules of the invention are also useful for identifying
XX	XX	a compound which alters ABCG5 activity level comprising contacting a cell
XX	XX	culture or mammal which have ABCG5 polypeptide with a compound and
XX	XX	measuring ABCG5 biological activity in the cell culture or in mammal,
XX	XX	where an increase or decrease in ABCG5 biological activity compared to
XX	XX	ABCG5 biological activity in a control cell culture or mammal not
XX	XX	contacted with the compound, identifies a compound that increases or
XX	XX	decreases ABCG5 activity respectively. The cell culture or mammal
XX	XX	comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
XX	XX	ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
XX	XX	polypeptide in a cell culture or mammal is also compared with that of a
XX	XX	second cell culture or mammal comprising a wild type ABCG5 polypeptide.
XX	XX	Stimulation of ABCG5 activity is useful for treating or preventing
XX	XX	hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
XX	XX	disease. The method of the invention is useful for increasing cholesterol
XX	XX	excretion and/or decreasing cholesterol adsorption. The present amino
XX	XX	acid sequence represents the mouse ABCG5 protein of the invention.
XX	XX	Sequence 652 AA;
XX	XX	Query Match 82.3%; Score 2738.5; DB 23; Length 652;
XX	XX	Best Local Similarity 80.1%; Pred. No. 1.2e-279;
XX	XX	Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1
QY	1	MGDLSILPFGSGMGLQVNRGSSQSSLEGAAPATAPED-HISGITLHAASYSHRRVPMWDDITS 59
Db	1	MGELPFLSPBPBARPHNRKSLSSLSQGSVYTERHSLSLGVLAVSYVSNRGPPWNNKS 60
QY	60	CRCQQTROIILKDVSLVYESGQIMCTILGSSGSKTTLDMAGSGRLRAGCTFLGEVYVNGRA 119
Db	61	COQKMKRDIILKDVSYLIESGGIMCTILGSSGSKTTLDMAGSGRLRAGCTFLGEVYVNGRA 120
QY	120	LRRQPDQCFEYVLOSDVFLSSLVYRETLRYTAMALCRSSADPFYNNKVEAVMTSLSH 180
Db	121	LRRQPDQCFEYVLOSDVFLSSLVYRETLRYTAMALCRSSADPFYNNKVEAVMTSLSH 180
QY	180	VADRLIGNYSLIGSTGERRRYSINAOLLOPKVWLPDEPTGLDCMTRANOIVLVLLAELA 239
Db	181	VADQMISYNGGSSGERRRYSINAOLLOPKVWLPDEPTGLDCMTRANOIVLVLLAELA 240
QY	240	RNRRIIVVLTTHQPSLEFQLEFDKTAIALSFGEILIFCGTFAEMIDFENDGYPCEPHSNPFD 299

Db 241 RRRIIVITVTHQPSRELFQHFQKIALITYGELVPGCTBEMLGFPPNCGIYCPHSHNPFD 3000

Qy 300 FYWDLTSVDTQSKEREIEFSKRVOMIESAYKSAICHRTLNKINIERMKHLKTLPMVPEPKT 3598

Db 301 FYMDLTSVDTQSKEREIEFTYKRVQMECAFKESEDYIHKILEMIERAFYKTLPIVPEPKT 3600

Qy 360 DSGCYGSKIGVLLRRTRNLVNRKLAIVTRLLQNLIMGLFLEFVLRYRSVTLKGAIDR 4198

Db 361 DPGMEKRLGVLLRRTRNLNRKQAVIMRLVQNLIMGLFLEFVLRYQNTLKGAVQDR 4200

Qy 420 VGLLYQVFGATPYTGMLANVNEFPVLRASVDSODGXYOKKQMLAVALLVLEFSVAT 4798

Db 421 VGLLYQVFGATPYTGMLANVNEFPMLRAVSDESODGXYHKQMLAVALLVLEFSVAT 4800

Qy 480 MFSVVCYWTGLHPEVARGYFSALLAPHILIGFELTVLGIYQNPINVSVALLSI 5398

Db 481 VFSVVCYWTGLGYEPVARGYFSALLAPHILIGFELTVLGIYQNPINVSVALLSI 5400

Qy 540 AGVLVSGFIRNIOEHPPIPKIISFTFOKTCSEILVYNEFYGLNFTGSSNVSTNPM 5998

Db 541 SGLIISGFIIRNIOEHPIDKILIGYFPOKYCCELVYNEFYGLNFTGSGNTSLNHPM 6000

Qy 600 CAGTQGIQFIEKTCPGATSRFTMNFLLYSFPAIVYIGIVFKRDLISR 651

Db 601 CATTQGVQFIEKTCPGATSRFTMNFLLIGFPAIVYIGIVFKRDLISR 652

RESULT 11

AAU96986 standard; Protein; 652 AA.

AAU96986; 30-JUL-2002 (first entry)

Rat ABCG5 protein.

Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

Rat sp.

WC200227016-A2.

04-APR-2002.

25-SEP-2001: 2001WO-US29859.

25-SEP-2000: 2000US-235268P.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

(PATE/) PATEL S B.

(DEAN/) DEAN M.

Patel SB, Dean M;

WPI: 2002-416483/44.

N-PDSB; ABR51686.

Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic acid encoding the polypeptide, useful for treating sitosterolemia, arteriosclerosis and heart diseases -

Example 3; Page 45; 66pp: English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal.

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity, respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention.

Sequence 652 AA:

Query Match 82.0%; Score 2727.5; DB 23; Length 652;
 Best Local Similarity 79.4%; Pred. No. 1.8e-278;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

QY 1 MGDLSLTPGSGMGLQVNRKSSQSSLECAPATAPAP-HSLGILHASYSHRYRPMWDITSC 59
 DB 1 MGEPLPSPGARGHNNKSSQSSLEGSVTGSEARHSIGLVNVSFVSNRYGPMWNITSC 60
 QY 60 CROQWTRQILNDVSLVSGOIMCTLGSSGSKTLLDAMSGRGRACFTGEEVYVNGRA 119
 DB 61 COQKDKRILNDVSLVSGOIMCTLGSSGSKTLLDAMSGRGRACFTGEEVYVNGRA 120
 QY 120 LRREQPODCFSYVLSQDITLSSLVRETLHYTALAIRNGNGSFQKVEAMALSTLSH 179
 DB 121 LRBDQPODCVSLQSDVFLSSLVRETLHYTALAIRNGNGSFQKVEAMALSTLSH 180
 QY 180 VADRLIGNSLGGISGTERRRVSTAAQLDQPKVLFDEPTTGDCMTANQIVLLVEIA 239
 DB 181 VADQMGVNYNGGSSIGSSGERRRVSIAQLDQPKVLMDEPTTGDCMTANQIVLLVEIA 240
 QY 240 RNRNIVYTLIHOPRESELPOLFKAILISGELIPGCTPAEMEDPDCGCPDEHSNPD 299
 DB 241 RNRNIVYTLIHOPRESELPOLFKAILISGELIPGCTPAEMEDPDCGCPDEHSNPD 300
 QY 300 FYMDLTVSDTOSKEREIEFTSKRVOMIESAYKSAICHKTLNIEHMKHLKTLPMVPEKTK 359
 DB 301 FYMDLTVSDTOSKEREIEFTSKRVOMIESAYKSAICHKTLNIEHMKHLKTLPMVPEKTK 360
 QY 360 DSPGFSLKLVLRVTNLRNKLAVTRLLQNLIMGLFLEFVLRVRSNVKGAIDR 419
 DB 361 NPGFCKRLGVLRVTNLRNKLAVTRLLQNLIMGLFLEFVLRVRSNVKGAIDR 420
 QY 420 VGLTYQFYGATPYTGMLNAVNLFPYLRAVSDQESODGLYOKMOMLALAVLHSPSYAT 479
 DB 421 VGLTYQFYGATPYTGMLNAVNLFPYLRAVSDQESODGLYOKMOMLALAVLHSPSYAT 480
 QY 480 MIFSSVCYWTGLHPEVARFGYSFSAALLAPHLIGFELTVLLGIYONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLHPEVARFGYSFSAALLAPHLIGFELTVLLGIYONPNIVNSVALLSI 540
 QY 540 AGVYVSGELNIQEMPIPKIISFTFQKCYCELVNVEYGLNFTGSSNSVYTNPM 599
 DB 541 SGLLGSGFIRNIEEMPIPKIISFTFQKCYCELVNVEYGLNFTGSSNSVYTNPM 600
 QY 600 CAFTQGIQFIKTCGATSRFTMNFLLIYSPALVILGIYVFKIRHLLISR 651
 DB 601 CSMTQGIQFIKTCGATSRFTMNFLLIYSPALVILGIYVFKIRHLLISR 652

RESULT 12
 AA096991
 ID AA096991 standard; Proteio; 408 AA.

AC AA096991;

XX 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R408X protein sequence.
 XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutelin.
 XX Homo sapiens.
 OS Synthetic.
 CC Key Location/Qualifiers
 FT Misc-difference 408
 FT /note= "Wild-type protein truncated at this position"

WO200227016-A2.

04-APR-2002.

25-SEP-2001; 2001WO-US29859.

25-SEP-2000; 2000US-235268P.

(USSH) US DEPT HEALTH & HUMAN SERVICES.
 (PATR/) PATEL S B.
 (DEAN/) DEAN M.

Patel SB, Dean M;

WPI: 2002-416483/44.

Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 nucleic acid encoding the polypeptide, useful for treating
 sitosterolemia, arteriosclerosis and heart diseases

Claim 10; Page -; 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette
 gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 predisposition for developing sitosterolemia, arteriosclerosis or heart
 disease. The molecules of the invention are also useful for identifying
 a compound which alters ABCG5 activity level comprising contacting a cell
 culture or mammal which have ABCG5 polypeptide with a compound and
 measuring ABCG5 biological activity in the cell culture or in mammal,
 where an increase or decrease in ABCG5 biological activity compared to
 ABCG5 biological activity in a control cell culture or mammal not
 contacted with the compound, identifies a compound that increases or
 decreases ABCG5 activity, respectively. The cell culture or mammal
 comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 polypeptide in a cell culture or mammal is also compared with that of a
 second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 stimulation of ABCG5 activity is useful for treating or preventing
 hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 disease. The method of the invention is useful for increasing cholesterol
 excretion and/or decreasing cholesterol adsorption. The present amino
 acid sequence represents the human ABCG5 mutant R408X protein of the
 invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.

Sequence 408 AA:

Query Match 62.6%; Score 2081; DB 23; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.7e-210;
 Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLQVNRKSSQSSLECAPATAPAPPHSLGILHASYSHRYRPMWDITSC 60

DB 1 MGDLSLTPGSGMGLQVNRKSSQSSLECAPATAPAPPHSLGILHASYSHRYRPMWDITSC 60

QY 61 RQWTRQILNDVSLVSGOIMCTLGSSGSKTLLDAMSGRGRACFTGEEVYVNGRA 120
 DB 61 RQWTRQILNDVSLVSGOIMCTLGSSGSKTLLDAMSGRGRACFTGEEVYVNGRA 120

QY 121 RREOFODCFSTVLAQSDTLSSLTAVRETLHTYATALLAIRKNGNSFOKKVEAVMAELTSHV 180
 DB 121 RREOFODCFSTVLAQSDTLSSLTAVRETLHTYATALLAIRKNGNSFOKKVEAVMAELTSHV 180
 QY 181 ADRLIGNYSLGISTGERRRVSIAAQLDPRKMLFDEPTTGDCMTANQIVLVLLVAR 240
 DB 181 ADRLIGNYSLGISTGERRRVSIAAQLDPRKMLFDEPTTGDCMTANQIVLVLLVAR 240
 QY 241 RNRIVLTIHOPRSELTPOLPKIALISFGELTFCGTAPAMLDFFNDCGPPCEHSNPPDF 300
 DB 241 RNRIVLTIHOPRSELTPOLPKIALISFGELTFCGTAPAMLDFFNDCGPPCEHSNPPDF 300
 QY 301 YMDLTVDYOSKEKEIEITSKRVMIESAYKSAICHHTLKNIEKMKHLKTLPMVPFKTD 360
 DB 301 YMDLTVDYOSKEKEIEITSKRVMIESAYKSAICHHTLKNIEKMKHLKTLPMVPFKTD 360
 QY 361 SPGVFSKLGVLRRVTNVLNKLAVITRLLQNLINGLFLFFVLAVR 408
 DB 361 SPGVFSKLGVLRRVTNVLNKLAVITRLLQNLINGLFLFFVLAVR 408

RESULT 13
 ID AAU96987 standard; Protein; 340 AA.
 AC AAU96987;
 DT 30-JUL-2002 (first entry)
 DE Hamster ABCG5 protein.
 KM Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 OS Cricetinae sp.
 PN NC0200227016-A2.
 PD 04-APR-2002.
 PF 25-SEP-2001; 2001MO-US29859.
 PR 25-SEP-2000; 2000US-235268P.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 DR N-PSDB; ABR51687.
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS Example 3; Page 46; 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal. CC where an increase or decrease in ABCG5 biological activity compared to CC ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or CC decreases ABCG5 activity respectively. The cell culture or mammal CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the hamster ABCG5 protein of the invention.
 SQ Sequence 340 AA:
 Query Match 41.7%; Score 1387.5; DB 23; Length 340;
 Best Local Similarity 78.6%; Pred. No. 2,2-137;
 Matches 264; Conservative 39; Mismatches 32; Indels 1; Gaps 1;
 QY 98 AMSGRDAGTFLGEVYVNGRALRREOFODCFSTVLAQSDTLSSLTAVRETLHTYATALLAIR 157
 DB 1 AISGRRLRTGTLEGEVFNNGRELRRDQFODCFSTVLAQSDVFLSLVRETLHTYATALLAIR 60
 QY 158 RGNFGSFOKKVEAVMAELTSHVADRILGNTSLGISTGERRRVSIAAQLDPRKMLFD 217
 DB 61 SSSDFYDKKVEAVMAELTSHVADRILGNTSLGISTGERRRVSIAAQLDPRKIMFD 120
 QY 218 EPTTGDCMTANQIVLVLLVARRRRIYVLTIHOPRSELTPOLPKIALISFGELTFCGTAP 277
 DB 121 EPTTGDCMTANQIVLVLLVARRRRIYVLTIHOPRSELTPOLPKIALISFGELTFCGTAP 180
 QY 278 AEMLDFFNDCGPPCEHSNPPDFYMDLTVDYOSKEKEIEITSKRVMIESAYKSAICHK 337
 DB 181 EEMLDFFNDCGPPCEHSNPPDFYMDLTVDYOSKEKEIEITSKRVMIESAYKSAICHK 240
 QY 338 TKNIEKMKHLKTLPMVPFKTOSPGVFSKLGVLRRVTNVLNKLAVITRLLQNLING 397
 DB 241 TKNIEKMKHLKTLPMVPFKTOSPGVFSKLGVLRRVTNVLNKLAVITRLLQNLING 300
 QY 398 LFLFFVLAVRVSNTLKGAIODRVGLTKQFGATPYT 433
 DB 301 LFLFFVLAVRVSNTLKGAIODRVGLTQFGATPYT 335

RESULT 14
 ID AAU96988 standard; Protein; 243 AA.
 AC AAU96988;
 DT 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R243X protein sequence.
 KM Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutain.
 OS Homo sapiens.
 PN NC0200227016-A2.
 PD 04-APR-2002.
 PF 25-SEP-2001; 2001MO-US29859.
 PR 25-SEP-2000; 2000US-235268P.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;
 Key Location/Qualifiers
 FT MISC-difference 243
 FT /note- "Wild-type protein truncated at this position"

XX WPI: 2002-416483/44.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 XX Claim 13; Page -: 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R243X protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX
 XX Sequence 243 AA:

Query Match 37.1%; Score 1234; DB 23; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.1e-121;
 Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCDLSLPFGSGNGIQLVNRGSSQSLGAPATAPEPHSLGILHASYSVSHRPVPMMDITSC 60
 DB 1 MCDLSLPFGSGNGIQLVNRGSSQSLGAPATAPEPHSLGILHASYSVSHRPVPMMDITSC 60
 QY 61 RQOMTRQILKDVSLVYESQIMCIIIGSSGSGKTTLLDAMSGRLGAGTFLGEVYNGRAL 120
 DB 61 RQOMTRQILKDVSLVYESQIMCIIIGSSGSGKTTLLDAMSGRLGAGTFLGEVYNGRAL 120
 QY 121 RREQFODCFYVLOSSTLLSLVRETHYTLALAIRGNPSSFOKVEAVNAEISLSHY 180
 DB 121 RREQFODCFYVLOSSTLLSLVRETHYTLALAIRGNPSSFOKVEAVNAEISLSHY 180
 QY 181 ADRLIGNVSLGISTGERRRVSIQAQLQDPVLMFDEPTTGLDCTANQIVYLVEIAR 240
 DB 181 ADRLIGNVSLGISTGERRRVSIQAQLQDPVLMFDEPTTGLDCTANQIVYLVEIAR 240
 QY 241 RNR 243
 DB 241 RNR 243

RESULT 15
 ID AAB41856
 ID AAB41856 standard; Protein; 144 AA.

AC AAB41856;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF1620 polypeptide sequence SEQ ID NO:3240.

KM Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KM vulnereary; antiparkinsonian; antiparkinsonian; nootropic; neuroprotective;

KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antineoplastic;
 KM antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M.

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76065.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2444; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnereary;
 CC antiparkinsonian; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 144 AA;

Query Match 21.7%; Score 722; DB 21; Length 144;
 Best Local Similarity 99.3%; Pred. No. 1.1e-67;

Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 357 KTKDSPGVFSKLGVLRRVTNLRNKLAVITRLQNLIMGLFLFPLRVSNVLCGAI 416
 DB 1 KTKDSPGVFSKLGVLRRVTNLRNKLAVITRLQNLIMGLFLFPLRVSNVLCGAI 60
 QY 417 ODRVGLLYQFVGATPYTGMLNANVLFVLAIVASQESODGILYQKQMMALVALVLPFSV 476
 |||

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rag

Page 13

Db 61 QDRVGLXQFYGAAPPYTGMNAVNLEPVLRAVSDQESODGLYOKWOMMLAYALHVLPSV 120

Oy 477 VATMIFSSVCYTWTGLHPEVARFG 500
|||||

Db 121 VATMIFSSVCYTWTGLHPEVARLG 144
|||||

Search completed: June 11, 2003, 09:12:15
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 11, 2003, 09:04:42 ; Search time 15 Seconds

(Without alignments)
1800.073 Million cell updates/sec

Database: US-09-989-981a-6

Search score: 3326

Sequence: 1 MGDLSLTPGSGMGLQVNRG.....PALVILGIVFKIRDLISR 651

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	3326	100.0	651	ABG5_HUMAN
2	2738.5	65.2	1	ABG5_MOUSE
3	2727.5	65.2	1	ABG5_MOUSE
4	698	21.0	673	ABG8_MOUSE
5	697	21.0	673	ABG8_MOUSE
6	691	20.8	672	ABG8_MOUSE
7	676.5	20.3	655	ABG2_HUMAN
8	627	18.9	1294	TOH5_YEAST
9	623	18.7	677	WHIT_YEAST
10	621	18.7	1049	ADP1_YEAST
11	607.5	18.3	695	WHIT_MOUSE
12	602.5	18.1	687	WHIT_MOUSE
13	596.5	17.9	678	ABG1_HUMAN
14	591	17.8	679	WHIT_MOUSE
15	578.5	17.4	646	ABG4_HUMAN
16	573	17.2	709	WHIT_MOUSE
17	569.5	17.1	666	ABG1_MOUSE
18	561.5	16.9	598	YPC3_MOUSE
19	547	16.4	666	YPC3_MOUSE
20	517.5	15.6	610	YPC3_MOUSE
21	454.5	13.7	675	BROW_MOUSE
22	435	13.7	675	BROW_MOUSE
23	427	12.8	1499	CDR3_MOUSE
24	406.5	12.2	1333	YPC3_MOUSE
25	403.5	12.1	1564	YPC3_MOUSE
26	399.5	12.0	1530	BFR1_MOUSE
27	398	11.8	1501	CDR3_MOUSE
28	392.5	11.8	1529	CDR3_MOUSE
29	391	11.8	1529	CDR3_MOUSE
30	383.5	11.5	650	ABG3_MOUSE
31	382	11.5	1501	CDR1_MOUSE
32	382	11.5	1511	CDR1_MOUSE
33	374	11.2	1511	PDR5_MOUSE

ALIGNMENTS

RESULT 1	ID	ABG5_HUMAN	STANDARD	PRT	651 AA
AC	Q9H222	ABG5_HUMAN	STANDARD	PRT	651 AA
DT	15-JUN-2002	(Rel. 41, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	ATP-binding cassette, sub-family G, member 5 (Sterol-1).				
GN	ABG5.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. AND VARIANT GLD-604.				
RC	TISSUE=Liver;				
RA	Medline-20553648; PubMed-11099417;				
RA	Berge R.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,				
RA	Willeverich J., Shan B., Barnes R., Hobbs H.H.,				
RA	Accumulation of dietary cholesterol in the liver of mice with				
RA	mutations in the ABCG5 and ABCG8 genes causes				
RA	Science 290:1771-1775(2000).				
RL	Science 290:1771-1775(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A. AND VARIANTS SITOSTEROLEMA H-389; H-419 AND P-419,				
RC	TISSUE=Liver;				
RA	Medline-20578753; PubMed-11138003;				
RA	Lee M.-H., Lu K., Hazard S., Yu H., Shulman S., Hidaka H., Kojima H.,				
RA	Dean M., Patel S.B.,				
RA	Identification of a gene, ABCG5, important in the regulation of				
RA	dietary cholesterol absorption.				
RA	Nat. Genet. 27:79-83(2001).				
RN	[3]				
RP	REVIEW.				
RA	Medline-21474438; PubMed-11590207;				
RA	Schmitz G., Langmann T., Helmerl S.,				
RA	Role of ABCG1 and other ABCG family members in lipid metabolism.				
RA	J. Lipid Res. 42:1513-1520(2001).				
RN	[4]				
RP	VARIANTS SITOSTEROLEMA Q-146; H-389; P-419; H-419 AND S-550, AND				
RA	Medline-21344600; PubMed-11453359;				
RA	Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,				
RA	Dean M., Patel S.B.,				
RA	Two genes that map to the STRL locus cause sitosterolemia: genomic				
RA	structure and spectrum of mutations involving sterol-1 and				
RA	sterolin-2, encoded by ABCG5 and ABCG8, respectively.				
RA	Am. J. Hum. Genet. 69:278-290(2001).				
CC	-1- FUNCTION: Transporter that appears to play an indispensable role				
CC	in the selective transport of the dietary cholesterol in and out				
CC	of the enterocytes and in the selective sterol excretion by the				
CC	liver into bile.				
CC	-1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to				

RT mutations in adjacent ABC transporters. ;
 RL Science 290:1771-1775(2000).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/reinoic X receptor (LXR/RXR) pathway.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF312713; AAC53097.1; -
 DR MGD: MG11351659; Abc95.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran. 1.
 DR ProDom: PD000006; ABC_transporter. 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR KMW: ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385 1 (POTENTIAL).
 FT TRANSMEM 386 406 1 (POTENTIAL).
 FT TRANSMEM 407 422 2 (POTENTIAL).
 FT TRANSMEM 423 443 2 (POTENTIAL).
 FT TRANSMEM 444 463 3 (POTENTIAL).
 FT TRANSMEM 464 484 3 (POTENTIAL).
 FT TRANSMEM 485 504 4 (POTENTIAL).
 FT TRANSMEM 505 525 4 (POTENTIAL).
 FT TRANSMEM 526 529 5 (POTENTIAL).
 FT TRANSMEM 530 550 5 (POTENTIAL).
 FT TRANSMEM 551 622 6 (POTENTIAL).
 FT TRANSMEM 623 643 6 (POTENTIAL).
 FT DOMAIN 644 652 6 (POTENTIAL).
 FT NP_BIND 87 94 6 (POTENTIAL).
 FT CARBOHYD 410 410 6 (POTENTIAL).
 FT CARBOHYD 585 585 6 (POTENTIAL).
 FT CARBOHYD 592 592 6 (POTENTIAL).
 FT CARBOHYD 592 592 6 (POTENTIAL).
 SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;
 Query Match 82.3%; Score 2738.5; DB 1; Length 652;
 Best Local Similarity 80.1%; Pred. No. 1.3e-183;
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;
 OY 1 MGDLSLTGSGMGLOVNRGSSSSLEGATAPAP-HSGILHASTSVSHRRPMDITS 59
 DB 1 MELPELSPGAPGAPINGSLSSLEQGSVGTGEANHSGLVHVSYSNRYGPMWNIRS 60
 OY 60 CROOFTQILKDVSLTVESGOINCLIGSSGSGKTTLLDMGSLGAGFLFVYVNGRA 119
 DB 61 CQKMDROILKDVSLTVESGOINCLIGSSGSGKTTLLDMGSLGAGFLFVYVNGRA 120
 OY 120 LRRDFQDFSVYVLOSLLSLTVRETLHTALLAIRNGSGSFQKTEAVAAELSLSH 179
 DB 121 LRRDFQDFSVYVLOSLLSLTVRETLHTALLAIRNGSGSFQKTEAVAAELSLSH 180
 OY 180 VADRLGNISLIGISIGERRRVSIAAQLQDPKVALFDEPTTGLDQMTANOQIVILLVELA 239
 DB 181 VADRLGNISLIGISIGERRRVSIAAQLQDPKVALFDEPTTGLDQMTANOQIVILLVELA 240

OY 240 RRRRIYVLTTHOPRSELEFQDKIAISFELFCGTPALMDPENCDCPCPEHSNPD 299
 DB 241 RRRRIYVLTTHOPRSELEFQDKIAISFELFCGTPALMDPENCDCPCPEHSNPD 300
 OY 300 FYMDLTSVDTQSKERELETSKRYQMTESAAKKAICHKTLKNTKMKHLLTLPMPVPEKTK 359
 DB 301 FYMDLTSVDTQSKERELETSKRYQMTESAAKKAICHKTLKNTKMKHLLTLPMPVPEKTK 360
 OY 360 DSGVSKGLVLLRRTYRNRYKKAIVITRLONLNGLEFLFVLRNSVULKAIQDR 419
 DB 361 DSGVSKGLVLLRRTYRNRYKKAIVITRLONLNGLEFLFVLRNSVULKAIQDR 420
 OY 420 VGLIYFVGATPYTGMNAVNLPVLAASVDSOSDGLYOKOMMLAYALHYLFESVAT 479
 DB 421 VGLIYFVGATPYTGMNAVNLPVLAASVDSOSDGLYOKOMMLAYALHYLFESVAT 480
 OY 480 MIFSSVCYWTGLHPEVARGFSAALLAPHLIGELTTLVLAGIVONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLHPEVARGFSAALLAPHLIGELTTLVLAGIVONPNIVNSVALLSI 540
 OY 540 AGVLSGFLRNQEMPIPKIISYTFQKCYSELVNEFYGLNFTCGSSNVSVTPNPM 559
 DB 541 SGLIISGFLRNQEMPIPKIISYTFQKCYSELVNEFYGLNFTCGSSNVSVTPNPM 600
 OY 600 CAFTGCIQFIEKTCPCATSRFTNMNFIIVSEIPALVILGIVFKIRDHLSR 651
 DB 601 CAFTGCIQFIEKTCPCATSRFTNMNFIIVSEIPALVILGIVFKIRDHLSR 652
 RESULT 3
 ID ABG5_RAT STANDARD; PRT; 652 AA.
 AC 099PE7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN ABG5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Granulata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_Taxid-10116;
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Small Intestine;
 RX MEDLINE-20578753; PubMed-11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Dean M., Patel S.B.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RT Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption. ;
 RL Nat. Genet. 27:79-83(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF312714; AAC53098.2; -
 DR InterPro: IPR003593; AAA_ATPase.

InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding: Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385
 FT TRANSMEM 386 406
 FT DOMAIN 407 422
 FT TRANSMEM 423 443
 FT DOMAIN 444 463
 FT TRANSMEM 464 484
 FT DOMAIN 485 504
 FT TRANSMEM 505 525
 FT DOMAIN 526 529
 FT TRANSMEM 530 550
 FT DOMAIN 551 624
 FT TRANSMEM 625 645
 FT DOMAIN 646 652
 FT NP_BIND 87 94
 FT CARBOHYD 585 585
 FT CARBOHYD 592 592
 SQ SEQUENCE 652 AA; 73342 MW; 4DA2FE2BAB0DD59 CRC64;
 Query Match 82.0%; Score 2727.5; DB 1; Length 652;
 Best Local Similarity 79.4%; Pred. No. 7.7e-183;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;
 1 MDGLSLFPGSGMGLVNGSSSLEGAATAPAP-HSGLIHASYSVSHRRPMDITS 59
 1 MGEFLFLEPGARGHNRNNGSSSLEGGVTSERHSAGLVNVSFVSNRGPWMNKS 60
 60 CROQMTROLKNDVSLYESGQIMCITIGSSGKTTLDAMSGRLRAGTFLEGVNCGA 119
 120 LRRBPPOCFVYVLDSTLSSLYRETLATALLAIRNGPSSQKVEAVMAELSLSH 179
 121 LRRDPPOCFVYVLDSTLSSLYRETLATALLAIRNGPSSQKVEAVMAELSLSH 180
 180 VADRLIGNTSLGISTGERRRYSIAQLQDPKVMFLDEPTTGLOCMANTAVILLVELA 239
 181 VADQMGNTNFGSISGERRRYSIAQLQDPKVMFLDEPTTGLOCMANTAVILLVELA 240
 240 RNRRIYVLTIHQPSRELFQLEFKIALISFGLIFCGTPEMLDFNDCGYPCEHSNPD 299
 241 RNRRIYVLTIHQPSRELFQLEFKIALISFGLIFCGTPEMLDFNDCGYPCEHSNPD 300
 300 FYMDLTVDYQSKREIEIKSKRVQMIKSAICHKTLKNIERKHLKTLPMVPEYTK 359
 301 FYMDLTVDYQSKREIEIKSKRVQMIKSAICHKTLKNIERKHLKTLPMVPEYTK 360
 360 DSPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNTIMGLFLEFVLARVSNVLRGAIDR 419
 361 NRPGRFCKLGYLRRVTRNLVNRKLAIVTRLLQNTIMGLFLEFVLARVSNVLRGAIDR 420
 420 VGLLQFGADPYTGMLNANVLPVLRVAVDSQSDGXYOKQMMALVALHLPESVAT 479
 421 VGLLQFGADPYTGMLNANVLPVLRVAVDSQSDGXYOKQMMALVALHLPESVAT 480
 480 MEESVCYWTGLHBEVAFRGFSALLAPHLIGELTLYLVLYGONRNIVNSVALLSI 539
 481 VLESSVCYWTGLHBEVAFRGFSALLAPHLIGELTLYLVLYGONRNIVNSVALLSI 540
 540 AGVLVSGFLNRNIQEMPIFKIISYTFQKCYSELLVNEFGLNFTGSSNVSTVTPM 599
 541 SGLLISGFLNRNIQEMPIFKIISYTFQKCYSELLVNEFGLNFTGSSNVSTVTPM 600
 600 CAFTGIGPIEKTGPGATSRFTMNLILYSFIPALVILGIYVERKIRDLISR 651
 601 CSMTGIGPIEKTGPGATSRFTMNLILYSFIPALVILGIYVERKIRDLISR 652

RESULT 4
 ID ABG8_MOUSE STANDARD; PRT; 673 AA.
 AC 09D8M0:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (SterolIn-2).
 GN ABG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A. (ISOPFORMS 1 AND 2).
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose I., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolIn-1 and
 RT sterolIn-2, encoded by ABG5 and ABG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOPFORM 1).
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shilagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Mikalido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Blake J., Boffelli D., Bojunga M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohitsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kvalterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABG5 or be tightly coupled to
 CC ABG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding, possibly mediated
 CC by the liver x receptor/retinoid x receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABG8 (WHITE)
 CC SUBFAMILY.

```

CC -!- CAUTION: Seems to have a defective ATP-binding region.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF324495; AAR64079.1; -
DR EMBL: AK004871; BAB23630.1; -
DR MGD: MGI:1914720; Abcg8.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transporter.1.
DR PROSITE: PS00211; ABC_TRANSPORTER.1.
DR GlycoProtein: Transmembrane; transport; Alternative splicing.
FT DOMAIN 1 413
FT TRANSSEM 414 434
FT DOMAIN 435 447
FT TRANSSEM 448 468
FT DOMAIN 469 496
FT TRANSSEM 497 517
FT DOMAIN 518 526
FT TRANSSEM 527 547
FT DOMAIN 548 569
FT TRANSSEM 570 590
FT DOMAIN 591 639
FT TRANSSEM 640 660
FT DOMAIN 661 673
FT CARBOHYD 677 679
FT VARSLIC 319 377
FT SEQUENCE 673 AA; 75995 MW; 78012611A5D2559 CRC64;

Query Match 21.08; Score 698; DB 1; Length 673;
Best Local Similarity 28.7%; Pred. No. 3,2e-41;
Matches 194; Conservative 133; Mismatches 264; Indels 84; Gaps 18;

11 GSGGLQVNRGSSSLSEAPATAPEPHSGLLHASYSHYR-----PKWD-I 57
14 GYLDADASQGLDQSL-----FSSSDNSLVFTYSGQSWTEYRDLTYQVDIASQVPMPEQL 69
58 TSCROOQMTROI-----LKDVSLYSGVSGIMCILLSSGGSTTTLLDNGRGRGCTF 109
70 AQKRIPIRSHSSSDSCSELGRINISFYKRSQOMALITSSCGRASILDVITGR-GHGGM 128
110 -LGEVTYNGRRLRREQPDFQDFSVYQSDTLSSLVYRETHYATLAI-RRGNPGSFQK 167
129 KSGQIWINQPTPTQVLRKCAVHVRHDDQLPLNYETELAFIAQMLPRTFSGAQDKR 188
168 VEAVMAELSHVADLIGNYSLGISTGERBRVSTAQQLQDPKWLFEPTTGLDCMT 227
189 VEDVIAELRLRQCANTRVGFTYRVGSGGERRRVSIGVOLAMPGLIILDEPTSGIDSTF 248
228 ANQIVLVLELARNRIIVLTIHQPSLELFQDLKAILSEFGLFCGTPAEMLDFENDC 267
249 AHNLVYTLRLKNGNLVILSLHQPISDFRLDVLVMTSGPIYILGAQAOVQYFTSI 308
288 GTPCPHNSPDPEDYMDLTSVDQSKERELETSKRVQMIESAVYKSA-----LCHTKLN 341
309 GHCPKPYSPNADFYVDLTSDIRSKEREVAVEKASIALFELEKVGDFDPMKKAKE 368
342 IERMKLKTLPMPFETKDS-----FGVESKLVLRARVRLNVRNKLAVITRLLQN 393
369 LMTSTHTVSLTL-----TQDPLDCGTAVALEGMEIQFSTLIRKQISNDFRDLPTLLINGSEA 424
394 LINGELFLFFVLVRNSVNLKGAIDQVRGLLYQVGAFTYGMNANVLPVLRAVSDQES 453
425 CLMSLIGFLYGHGAKOL--SPMDNALFLMGAIIPIRVLLIDVYSKCHSEKSMYYTEL 482
454 QDGLQKQOMLALVALVLPFSVVAITMIFSSVCYWTILGHPVARCGYFSALLAHLIG 513

DB 483 EDGLYTAGPYEFAKILGELPERCAVYIYAMPYMLNLRPVELF-----LL--HLL 534
QY 514 EEL-----TLVLGIVQNPPI-VNSVVALISLGVGSGFLNRNIDMPIPEKISTFT 566
DB 535 VMLVVFCCIRNALASANLPTFHMSSEFCNALYNSFYLTAGFMINDMLIVPAMISKLS 594
QY 567 FOKYCEILVYNEPYGLNFT--CGSSNVSTYTNMCAFTQGIQTEKTCPATISFTMNF 624
DB 595 FLRMCFSGLMQIYNGHLITTYQIGNFTFYSILGDM-----ISAMDLS 637
QY 625 LILVSPALVYLCI 639
DB 638 HPLVAYI--LIVIGI 650

RESULT 5
ABG8_HUMAN STANDARD; PRT; 673 AA.
ID ABG8_HUMAN
AC 09H221;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterol-12).
GN ABG8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A. VARIANTS SITOSTEROLEMA T-231; Q-263; R-574 AND
RP R-596. AND VARIANT C-54.
RX MEDLINE-20533648; PubMed-11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishn N.V., Schultz J.,
RT Kvitrovich P., Shan B., Barnes R., Hobbs H.H.;
RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RL mutations in adjacent ABC transporters.";
RL Science 290:1771-1775(2000).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS SITOSTEROLEMA H-184;
RP T-231; Q-263; H-405; P-501; S-543; P-572; E-574; R-574; R-596 AND
RP F-570 DEL, AND VARIANTS H-19; C-54; K-238; V-259; K-400; R-575 AND
RP A-632.
RC TISSUE-Liver;
RX MEDLINE-21344600; PubMed-11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Oue L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Silvastava A.K.,
RA Patel S.B.;
RT "Two genes that map to the SRSF locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterol-1 and
RL sterol-2, encoded by ABCG5 and ABCG8, respectively.";
RL Am. J. Hum. Genet. 69:278-290(2001).
[3]
RP REVIEW.
RX MEDLINE-21474438; PubMed-11590207;
RA Schmitz G., Langmann T., Helmerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL J. Lipid Res. 42:1513-1520(2001).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC produced by alternative splicing. Isoform 1 (shown here) and 2; are
CC detected in approximately 10% of the cDNA clones.
CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
CC in the small intestine and colon. Detectable in a wide variety of
CC human tissues.
CC -!- DISEASE: Defects in ABCG8 are a cause of sitosterolemia, also

```

known as phytosterolemia or shellfish sterolemia, a rare autosomal recessive disorder characterized by increased intestinal absorption of all sterols including cholesterol, plant and shellfish sterols, and decreased biliary excretion of dietary sterols into bile. Sitosterolemia patients have hypercholesterolemia, very high levels of plant sterols in the plasma, and frequently develop tendon and tuberous xanthomas, accelerated atherosclerosis and premature coronary artery disease.

-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.

-1- CAUTION: Seems to have a defective ATP-binding region.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AF320294; AAG40004.1; -
EMBL; AF324494; AAK84078.1; -
EMBL; AF351824; AAK84663.1; -
EMBL; AF351812; AAK84663.1; JOINED.
EMBL; AF351813; AAK84663.1; JOINED.
EMBL; AF351814; AAK84663.1; JOINED.
EMBL; AF351815; AAK84663.1; JOINED.
EMBL; AF351816; AAK84663.1; JOINED.
EMBL; AF351817; AAK84663.1; JOINED.
EMBL; AF351818; AAK84663.1; JOINED.
EMBL; AF351820; AAK84663.1; JOINED.
EMBL; AF351819; AAK84663.1; JOINED.
EMBL; AF351821; AAK84663.1; JOINED.
EMBL; AF351822; AAK84663.1; JOINED.
EMBL; AF351823; AAK84663.1; JOINED.
Gene; HGNC:13887; ABCG8.
MIM: 605460; -
MIM: 210250; -
InterPro: IPR003439; ABC_transporter.
Pfam: PF00005; ABC_tran; 1.
ProDom: PD000006; ABC_transporter; 1.
ProSITE: PS00211; ABC_TRANSPORTER; 1.
GlycoPhorin: Transmembrane; Transport; Alternative splicing;
Kw Polymorphism: Disease mutation.
DOMAIN 1 416
TRANSMEM 417 437
DOMAIN 438 447
TRANSMEM 448 468
DOMAIN 469 492
TRANSMEM 493 513
DOMAIN 514 531
TRANSMEM 532 552
DOMAIN 553 569
TRANSMEM 570 590
DOMAIN 591 639
TRANSMEM 640 660
DOMAIN 661 673
CARBOHD 619 619
VARSPIC 376 376
VARIANT 19 19
VARIANT 54 54
VARIANT 184 184
VARIANT 231 231
VARIANT 238 238
VARIANT 259 259
VARIANT 263 263
/FTid-VAR_012250.
Y->C.
/FTid-VAR_012251.
R->H (IN SITOSTEROLEMIA).
/FTid-VAR_012252.
P->T (IN SITOSTEROLEMIA).
/FTid-VAR_012253.
E->K.
/FTid-VAR_012254.
A->V.
/FTid-VAR_012255.
R->Q (IN SITOSTEROLEMIA).
/FTid-VAR_012256.

FT VARIANT 400 400 T->R.
FT /FTid-VAR_012257.
FT VARIANT 405 405 R->H (IN SITOSTEROLEMIA).
FT /FTid-VAR_012258.
FT VARIANT 501 501 L->P (IN SITOSTEROLEMIA).
FT /FTid-VAR_012259.
FT VARIANT 543 543 R->S (IN SITOSTEROLEMIA).
FT /FTid-VAR_012260.
FT VARIANT 570 570 MISSING (IN SITOSTEROLEMIA).
FT /FTid-VAR_012261.
FT VARIANT 572 572 L->P (IN SITOSTEROLEMIA).
FT /FTid-VAR_012262.
FT VARIANT 574 574 G->E (IN SITOSTEROLEMIA).
FT /FTid-VAR_012263.
FT VARIANT 574 574 G->R (IN SITOSTEROLEMIA).
FT /FTid-VAR_012264.
FT VARIANT 575 575 G->R.
FT /FTid-VAR_012265.
FT VARIANT 596 596 L->R (IN SITOSTEROLEMIA).
FT /FTid-VAR_012266.
FT VARIANT 632 632 V->A.
FT /FTid-VAR_012267.
SQ SEQUENCE 673 AA; 75678 MW; 594AFDID6CIBB50F CRC64;
Query Match 21.0%; Score 697; DB 1; Length 673;
Best Local Similarity 28.9%; Pred. No. 3.8e-41;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;
QY 8 TTGGSGGLQVNGSGSSLEGAPAT-APEPHSGILHASTVSRRV-PWMD-ITSCROOM 64
DB 16 TPDTSGLQDRLFSSSESDNSLYFTYSGQNTLEVRDLNQVLAQSPWFEOLQAKRM 75
QY 65 TQOI-----LKQVSLVESGOIMCITIGSSGSGCTTTLLDMSGGLGAGTF-LGEYV 115
DB 76 TSPSCNCELQIQTNSFVRKSGOMALITGSSGCGASLLDVYTG-GGSGTKSGQIWI 134
QY 116 NGRALRRQFQDCFSYVLOSSTLSLSTVRETHYATLAI-RGNGPSFQKVEAVMAE 174
DB 135 NGQSSPOLYRKCVANVHRQHNOLPLTYRETLAIQAKRLRTSQAQRKVEDVIAE 194
QY 175 LSLSHVADRLIGNYSLGISTGERRRVSIAAQLDPPKYMLEDFTTGIDCMTANOIVYL 234
DB 195 LRLRCQADRVGNMVMYRGISGERRRVSIGVOLLNMPGILIDEPSTGIDSTFANHLVMT 254
QY 225 LVELARRNRIYVLTTHQPSSELPQFEDKAILSPGELICGPAEMLEDFNDCGTPCP 294
DB 255 ISRLAKGNRLVLSLHQPSPDIFRLEPDLVLTMTSGPTIYLGAAOHVQYFTAIIGYPC 314
QY 295 SNRPFDYMDLTSVDFQSKREIETSKRVQMISAVKKAISCHTKLNIRMKHL----- 348
DB 315 SNRPADFYDLSIDRRSRQELATREKQSLAALF-----LEKVDLDDFLMK 362
QY 349 -----KTLPN-----VPFKTKDSPGVSKLGVLLRVTNMLVRNKLAVITRL 390
DB 363 AETKDLDEDTCESSVTFPLDNCPLSPRK-MGAVQOFTTLIRQISDNDFRLPLLHG 421
QY 391 LQNLMLGLELFEVLRVSNVAKGAIQ-----DRVLQLQFVATITTYGTLNANVLEPVL 446
DB 422 AEACLSMTIGFLYIG-----HGSIQSEMDTALPMIGALIPFNILDIYSKCYSER 475
QY 447 AVSDQSDGLYQKQOMMLAYALHVPESVATMIFSSCYVTLCLHPEVARF----- 499
DB 476 AMLYELLEDGLYTTQPYFFAKTIGELPEHCAYIITYGAPTIVLANLRGLODFLLHLLV 535
QY 500 -----GFSALLAPHLIGFELVLVLGIVQNDNIVNSVALLSINGVLVSGFL 549
DB 536 WLVECCRIMAAALALPTFHMASFFS-----NALVNSFYLAG-----GFM 577
QY 550 RNIOQMPILPKIISFTFOKXCSLIVNVEYGLAFTGSSNSVSTTN 597
DB 578 IINLSLMTVPAMISKVSFLKWCFGELMKIQPSRRTYKMPDGNITVAVS 625

RESULT 6
ABG8_RAT STANDARD: PRT: 672 AA.

AC P58428: 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterol-2).
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Silvastava A.K.,
RA Patel S.B.;
RT "Two genes that map to the SRSF locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterol-1 and
RT sterol-2, encoded by ABCG5 and ABCG8, respectively.";
RL Am. J. Hum. Genet. 69:278-290(2001).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -1- CAUTION: Seems to have a defective ATP-binding region.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE351785; AAK84831.1;
CC Interpro: IPR003593; AAA_ATPase.
CC Interpro: IPR003439; ABC_transporter.
CC Pfam: PF00005; ABC_tran. 1.
CC SMART: SM00382; AAA. 1.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC Glycoprotein: Transmembrane; Cytoplasmic (POTENTIAL).
CC FT DOMAIN 1 412
FT TRANSSEM 413 433
FT TRANSSEM 434 446
FT TRANSSEM 447 467
FT TRANSSEM 468 495
FT TRANSSEM 496 516
FT TRANSSEM 517 525
FT TRANSSEM 526 546
FT TRANSSEM 547 568
FT TRANSSEM 569 589
FT TRANSSEM 590 628
FT TRANSSEM 629 649
FT TRANSSEM 650 672
FT TRANSSEM 618 618
FT CARBOHYD 376
FT VARSPIC 376
FT SEQUENCE 672 AA: 75878 MW: 608897353FA4C36F CRC64;
Query Match 20.8%; Score 691; DB 1; Length 672;
Best local Similarity 28.3%; Pred. No. 9.8e-41;

Matches 189; Conservative 126; Mismatches 264; Indels 88; Gaps 15;

OY 23 SLECAPATAPEPHSIGILHASYSVSHRVR-----PW-----WDITSC 60
DB 21 SLDQSVSSSDNSLFTYTGSGSNLTLEBRLTYQVDMAQVPMPEQLAQKLPWRSGS 80
OY 61 RQMTROLKVSILYESGQIMCIGSSGSKTLLDMSRGLRAGTGLCEVYVNGRAL 120
DB 81 QDSMDLGI-RNLSFKVRSGOMALIGSACGRATLIDVITGRDGGKSKSQIMNGPS 139
OY 121 RREGFODFSYVLQSDTLTSLTYRETHYATLAIKRGNGSF-----QKKVYMAEL 175
DB 140 TPQLQKCVAHVRQDDLLPMLTYRETHYATLAIKRGNGSF-----PRTSQQRDRVEDVIAEL 195
OY 176 SLSHVADRLNIGSLGISTGERRRVSIAAQLQDPKVMLEDEPTTGIDCMTANQIVYL 235
DB 196 RLROCAVTRVGNVTVRGVSGGERRRVSIGVQLLNNPGLILDEPTSGIDSTFANIVRTL 255
OY 236 VELARRNRIVVLTTHORSELPFLPKAILISFELLFPGTAPALDFPNDGYPCEHS 295
DB 256 SRLAKGRNLTSLIHPKSDIFRLELDVLLMTSGPIYLGVAAHMQVFTSIGYPCPRS 315
OY 296 NPPEFYDLTSDVDTQSKRETEETSKRVOMIESAKKSA-----ICKTKLNIEMRHLK 349
DB 316 NPADFYDLTSDVDTQSKRETEETSKRVOMIESAKKSA-----ICKTKLNIEMRHLK 371
OY 350 TLPMPVPEKTKDS-----PGVFSKGLVLRVTRNLRNKLAVITRLQNLINGFL 401
DB 372 TYAVSQTLQDPTNGCTAELRPMIOQFTTLRLQINDFRLPLFTLHGACMLSLIG 431
OY 402 FETVLRVSNVILKAIQDRVCLLYPGATPTTGKLAANVLEPVLRAVSDQSGLYQKW 461
DB 432 FLYYGHADKPL--SEPMMAALLFMIGALIPFNVLIDVYSKHSRSLYELLEDGLYTAG 489
OY 462 QMMLVALVLPFFSVVATMIFSSVCYTWLGHPPVARGYSALLNPHLIGEL----- 516
DB 490 PYFAKVLGELPEHCATVITIGMPITWLNRP-----GPEFLHPLMLLVFCC 541
OY 517 -TLVLGIVQNPNI-VNSVVALISAGLVGSLRNIOEMPIFKIISYTFQKYSCEI 574
DB 542 RTMLAASAMLPFPHMSFCCNALNSFYLAGMINLMNIVPMISKMSFLRMKFSG 601
OY 575 LVYVEFGFLNFTGSSSVSTTNWCAFTGIGTETCGE--ATSRFNFLIYSFIP 632
DB 602 LMQIQFNGHYTTQIGMLTFESV-----PGDAVMTAMDLSNHLVATY 643
OY 633 ALVILGI 639
DB 644 -LVYIGI 649

RESULT 7
ABG2_HUMAN STANDARD: PRT: 655 AA.

AC Q9UNQ0: Q95374; Q9UNQ0; Q9B773;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-
DE binding cassette transporter) (Breast cancer resistance protein).
GN ABCG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99065313; PubMed=9850061;
RA Allikmets R., Schriml L.M., Hutchison A., Romano-Spica V., Dean M.,
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on
RT chromosome 4q22 that is involved in multidrug resistance.";
Cancer Res. 58:5337-5339(1998).

RN [12] SEQUENCE FROM N.A.
 RP TISSUE-Breast cancer;
 RC MEDLINE-99080071; PubMed-9861027;
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RT Ross D.D.;
 RT "A multidrug resistance transporter from human MCF-7 breast cancer
 RT cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
 RN [13] ERRATUM.
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
 RN [14] SEQUENCE FROM N.A.
 RP Rake K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,
 RA Sugimoto Y.;
 RT "Breast cancer resistance protein constitutes a 140-kDa complex as a
 RT homodimer";
 RL Submitted (Mar-2001) to the EMBL/GenBank/DBJ databases.
 RN [15] SEQUENCE OF 198-655 FROM N.A.
 RP TISSUE-Placenta;
 RC Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,
 RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RT Niimura K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
 RN [16] REVIEW.
 RP MEDLINE-21474438; PubMed-11590207;
 RA Schmitz G., Langmann T., Helmerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
 CC IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
 CC CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME
 CC RESISTANT TO MITOXANTHRONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY
 CC DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
 CC MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUPERFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: AF103796; AAD09188.1;
 DR EMBL: AF098951; AAC97367.1;
 DR EMBL: AB056867; BAB39212.1;
 DR EMBL: AK002040; BAA92050.1;
 DR Gene: HGNC:74; ABCG2.
 DR MIM: 603756;
 DR InterPro: IPR0003593; AAA_ATPase.
 DR InterPro: IPR0003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transportr.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
 DR ATP-binding; Transmembrane; Transport.
 DR DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 396 416 POTENTIAL.
 FT TRANSMEM 417 428 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 429 449 POTENTIAL.
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 478 498 POTENTIAL.
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 507 527 POTENTIAL.
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 536 556 POTENTIAL.
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 80 87 ATP (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 24 24 Y -> A (IN REF. 2 AND 4).
 FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
 FT CONFLICT 208 208 F -> S (IN REF. 1).
 FT CONFLICT 315 316 MISSING (IN REF. 5).
 FT CONFLICT 316 316 R -> T (IN REF. 2).
 FT CONFLICT 482 482
 SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CE0 CRC64;

 Query Match 20.3%; Score 676.5; DB 1; Length 655;
 Best Local Similarity 29.0%; Pred. No. 9,8e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

 QY 21 SOSLEGAPATAP---EPHSGLHNSYSVSRVRRWMDITSGRQGTROIIDKDYLYE 77
 DB 13 SGNRTGFPATYINDLKATBEAVLSFNHCIRVKLSGFLPCRKEVEILSNINGIK 72
 QY 78 SQOIMCITGSSSGRTLLDAMSGRLGRATFGEYVYNGRALRREOFQDCFSYVLSQDT 137
 DB 73 PG-LNAILDPTGGKSSLLDVLAAKRDPSG-LSGCVLLNG-APRANPCNSGYVQDDV 129
 QY 138 LLSLTVRETLTYTLLAIRGNPG-SFOKVEAYVAEISLHVADRLGNTSLGISTG 196
 DB 130 VNGTLVRENLOFSAALRLATMTNHEKERNIRYEEIGLDKVDKSGTOFINGVSG 189
 QY 197 EERRVYSNOLLQDPRVLMFDEPTGLOMTQNOVIVLYELARRNRYVLTIRPRSL 256
 DB 190 EKRKSIGMELTDDSTSLDEPTGDSANMAVLLLRKSKGGRITTFIHOPRSI 249
 QY 257 FQLPDKIALISFGLIFGTPAEMLDFFNDGCPCEPHSNPDEFYMDLTSVPTQ----SK 313
 DB 250 FKLFDLSLTLASGRLMFHPGPAQEGALGYESAGYHEANNNPADFLDIINGSTVALNR 309
 QY 313 EEE-----IETSKR-----VQMBISAYKSAICHT-----LKNIERKHLKTLPMVP 356
 DB 310 EEDFKATELIEPSKODKPLIEKIAEIVYNSSPYKETKAKELHQLSGEKKKITYEKEISY 369
 QY 357 KTKDSPGVSKLGVLLRVTNRLVNRKLAIVTIRLLONIMGLF--LEFVLRYRNVNLKG 414
 DB 370 TT-----SCHOLRWYSKSKFKNLNPQASIAQIIVYVLLGLVIGAITFGKLNDS----- 421
 QY 415 AIQDRVGLLYQFVGATPYTGMNAVNLPEVYLAASVDSODGLYQKQOMMLAYAL-HYLP 473
 DB 422 GIONRAGLYF-FLTNQCFSSYSAVELVEYVEKFLIHYISGYRVSSYPIGLKLSIDL 480
 QY 474 FSVVATMTESVVCYTGLIHPVNAFGFSNALLAPHIGELTIVLYLGIONPRTYNSV 533
 DB 481 MRMLDSILFTCIYFMGLKPRADAFYMMETIM---WVAYSASSMALAIAGOSVSVSA 537
 QY 534 VALLSIAGV--LVGSGFLRNIOEMPIPKIISYFFQKCGSEILVNVNEGTGNTGSSSN 591
 DB 538 TLIMTICVFMFMISGLLVNLTITASMLSMQYTSIPRTGTTALOHNEFLQNFPCPG--- 594
 QY 592 VSVTTPKCAFTQGIQTEKTCPG 615
 DB 595 LNAIGNNPCNYA-----TCTG 610

 RESULT 8
 YOH5_YEAST
 ID YOH5_YEAST STANDARD; PRT; 1294 AA.

AC 008234; 008233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97321807; PubMed=9178509;
 RA Tzeremia M., Katsoulou C., Alexandraki D.;
 RT Sequence analysis of a 33.2 kb segment from the left arm of yeast
 RT chromosome XV reveals eight known genes and ten new open reading
 RT frames including homologues of ABC transporters, Inositol
 RT phosphatases and human expressed sequence tags.
 RL Yeast 13:583-589(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 DR EMBL: 274817; CA989065.1; -
 DR EMBL: 274816; CA99084.1; -
 DR SGD: S0005435; YOL075C.
 DR InterPro: IPR003593; AAA_Atpase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 FT TRANSMEM 376
 FT TRANSMEM 496
 FT TRANSMEM 531
 FT TRANSMEM 605
 FT TRANSMEM 1039
 FT TRANSMEM 1121
 FT TRANSMEM 1267
 FT NP_BIND 62
 FT NP_BIND 727
 FT CARBOHYD 41
 FT CARBOHYD 86
 FT CARBOHYD 101
 FT CARBOHYD 151
 FT CARBOHYD 151
 FT CARBOHYD 341
 FT CARBOHYD 349
 FT CARBOHYD 371
 FT CARBOHYD 528
 FT CARBOHYD 528
 FT CARBOHYD 983
 FT CARBOHYD 1062
 FT CARBOHYD 1062
 FT CARBOHYD 1294
 FT SEQUENCE 145157
 FT MW: C555500A45E9284E CRC64;
 SQ
 Query Match 18.9%; Score 627; DB 1; Length 1294;
 Best Local Similarity 31.7%; Pred. No. 6, 7e-36;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

OY 65 TROIADVSLVYSGQIMCIGSSGKTTLLDANSGRGRAGTFL-----GEVYNGRA 119
 DB 706 TREIIOVNAIFKPMINAKIPSSGKSSLLNISGR-KSSVAKFPTSGSIMENDIO 764
 OY 120 LRREQFOCFSTVLAQSDT-LSSLTVEVRLTATLAIIRNGPFGFOKKEVAVMAELIS 178
 DB 765 VSELFKNVCSVSODDHLALATLVKTKYAAALRLHHLTEAREMENTDLINSIGLK 824

OY 179 HVADRLIGNSLGISTGERRRYSIAQLQDPRKVMLEFDEPTGLDCMTANQIVLVLEL 238
 DB 825 HCENNITIGNEFVIGSGEKRRRYTMGQVLNDPILLDLDPSTGLDSFATLLEIKL 884
 OY 239 AR-RNRIVLTIHQPRSELFOLEDKIAIIS-FGELIPGTPAELEDFNOCGPCPEHSN 296
 DB 885 CRGQKTIITTIHQPRSELFOLEDKIAIIS-FGELIPGTPAELEDFNOCGPCPEHSN 944
 OY 297 PEDFYMDLISVDQSKREIETSKRVQMISAVKKSACIKTKLNIEMRMHKLTPMPVF 356
 DB 945 VADFFLDLISVNTQNEQNEISSRAREVKEIISAK-----ANNON-ESLSTPISSEK 994
 OY 357 KTDSPVEFKGLVLRVTRNV-----RNLAVITRLQNLIMGLFL 401
 DB 995 QOYSOSEFFTEYEEFVKR-PANLVAYIVVKRQFTTTRRSFDSIMARIAIQIPGLVIFA 1053
 OY 402 FVILRVRSVILKGAIDPRVGLYQFVATPYTGMLNANVLPVLRVANSDDSDGLQKW 461
 DB 1054 LFFAPRKHNT--SISRDLADEST-ALTFVGLGMLACTPERDYFTEYNDNVGIA 1110
 OY 462 QMMLAYALHVPFSVATMTFSSVCYWTGLIHEVARFGYSAALLAPHLI---GEFLTL 518
 DB 1111 PFLAYVTLLEPLSALASVAVFTVLACL-PRTA--GNFPATVCSFVTCGHEGLI 1167
 OY 519 VLLGIQVQNR-YNVSVALSLAGVLSGFLRNIDQMPLEPKISFTFOKCSSELVV 577
 DB 1168 MWTFFERPEFVNCIISTILSTQMSGLMSL-----GMSRVLGFNTLPVGYTSMIIN 1223
 OY 578 NEFYG-LNFTC--GSSNVSVTNPMACTOG 605
 DB 1224 FAFPNKILKCEGKNSDST-----CEFANG 1250

RESULT 9
 ID WHIT LUCCU STANDARD; PRT; 677 AA.
 AC 005360;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Oestroidea; Calliphoridae; Luciflia.
 OX NCBI_TaxID=7375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97087158; PubMed=8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.;
 RT "The structure, sequence and developmental pattern of expression of
 RT the white gene in the blowfly *Lucilia cuprina*."
 RL Insect Mol. Biol. 5:251-260(1996).
 RP [2]
 RP SEQUENCE OF 490-584 FROM N.A.
 RX MEDLINE=90264941; PubMed=1971656;
 RA Elizur A., Vacek A.T., Howells A.J.;
 RT "Cloning and characterization of the white and topaz eye color genes
 RT from the sheep blowfly *Lucilia cuprina*."
 RL J. Mol. Evol. 30:347-358(1990).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: 038899; AAA82057.1; -
 CC EMBL: X53265; CAA37365.1; -
 DR InterPro: IPR003593; AAA_Atpase.
 DR InterPro: IPR003439; AAA_transport.
 DR InterPro: IPR005284; pigment_permease.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRFAMs: TIGR00955; 3a01204; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR Pigment: ATP-binding; Transmembrane; Transport.
 FT NP_BIND 119 126 ATP (POTENTIAL).
 FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 534 554 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 647 667 POTENTIAL.
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;

Query Match 18.7%; Score 623; DB 1; Length 677;

Best Local Similarity 27.0%; Pred. No. 5,5e-36; Matches 188; Conservative 144; Mismatches 260; Indels 104; Gaps 19;

8 TPGGMGLOVNGSSSLEGAPATAPPEPHSLGILNASVSHRVRPW-----DITSC 60
 27 TPG-----TLEASAINSGFSKSYGSIVSNESAEKLYTSMCNLDVGEVHOP 73
 61 RQW-----TROIKDVSLVESQIQCILGSSGSGKTTLLDAMS 100
 74 GSNMQLVNRKVCNEHRIKPKRKHKVCAVPELLAVMSSGAGKTTLLNALA 133
 101 GRIGAGTFLGEVY---NGRALREPODCEVYLSISSLTVPRELYTALLAR 157
 134 FRSRAR-GVOISPSYRMLNGHYDAKEMQARCAIVQDDLFGLSFARHLLFQATVRP 192
 158 RGNPSFG-KRYEAVMAELSLSHVADRIG-NYSLGISTGERRRYINAOLOPKVM 215
 193 RTMTQKRLQRYDQYIQLSLIKCNITIGVGRKGLSGGERKKRLAFSEALTPPLLI 252
 216 FDEPTGIDCMANIVVLVELARNRIVYITIQPSSELPOLDKIALISFGLIFCG 275
 233 CDEPTSGIDSEMAASVOYLKRLSGRGKTVLITIQPSSELPOLDKIALISFGLIFCG 312
 276 TPAMLDPEFNDGCPPEHNPEDFYMDLTSVDTOSKREIETSKRYQMIESAAYKKAIC 335
 313 TPVEAVDFEFLGACCPINYNPADPYOVLAV---VPGREISRRIRISICDNFAYKYS 369
 336 HKTAKNIERMKHKLTPVAPPKT---KDSGV-----FSKLGVLRLRYTRNLVRNKL 384
 370 REMENQFQK-----IAAKTDGLQKDDETTLIKKASWTFQFRAIMRMSISTLKEPL 420
 385 AVITRLNLNLGLFL-LFVYLARNSVNLKAIQORVGLIYQVCAPTPTGLANVNLFP 443
 421 LVKVALIOTTYAAVLIGLIFLNMPTQY---GVAMINGAIFLEFLNMFTQNFVAIVNFT 477
 444 VLRASDQSDGLYOKWQMLAYALHVLFPESVATMTIFFSCYCYTLGHPVAREGES 503
 478 SELPVEFMETSRILRCOTYFLGKRLAELPLFIVPFLFIAIAYMIGLRBIT---HFL 534
 504 AALLAPHLIGELT---LVLLGIYONPNIIVNSVALLSIAGVLSSGFLRNIOEMPIPK 560
 535 SALALVLYVANVSIFGLISCASSTSMALSVGPHLIPILFLFGVFL-NGSVPVYFK 593
 561 IISYTFQKYSSEILVNEFYGL---NFTGSSNSVNTNMCACTOGIOFIEKCP--G 615
 594 WLSTFSMRIRANEGILINQMDVQGEITCTSTN-----TCPSG 634
 616 ATSRTNFLIYSPALVITIGIVFERIDHLISR 651

DB 635 XYXLETLNFRDXFTFRLYLILILIFRIAGYVAK 670
 RESULT 10
 ADPL_YEAST STANDARD; PRT; 1049 AA.
 ID ADPL_YEAST
 AC P25371;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Probable ATP-dependent permease precursor.
 GN ADPL OR YCR011C OR YCR105.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92160395; PubMed-1789009;
 RA Purnelle B., Skala J., Goffeau A.;
 RT "The product of the YCR105 gene located on the chromosome III from
 RT Saccharomyces cerevisiae presents homologues to ATP-dependent
 RT permeases.";
 RL Yeast 7:867-872(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92327849; PubMed-1626432;
 RA Skala J., Purnelle B., Goffeau A.;
 RT "The complete sequence of a 10.8 kb segment distal of SUF2 on the
 RT right arm of chromosome III from Saccharomyces cerevisiae reveals
 RT seven open reading frames including the YRS161, ADP1 and PGK genes";
 RL Yeast 8:409-417(1992).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation,
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

OY 388 TRILQWLMGLFLLFVLRVRSVNLGALIDRVGL-----LYQVGCAPPTTGMNLAVNL 441
 DB 442 VRLLOTATNAVATLI-----GSIFYGQVLDGCVNMNINSLFLFLNMTFQVAVINV 493
 OY 442 FPLRLRVSDQESDDGIYOKKQMLAALAHLPFSVAVATMFSSVCWTLCLEHVARFGY 501
 DB 494 FSAELPFLERKRSRLRYRDVYPLGRTIETLPLEFIVPVPFTSITPMICL-----RTG- 547
 OY 502 FSNALLAPHLIGELFLVLLGIYQNP-----IVNSVVMALSLA-----GVLVGSG 547
 DB 548 -----ATHYL-----TLFLYITLVANVSTSGVYILCSASSISMAISVGPVPVPLIFGG 598
 OY 548 FLRNIGDMPILPEFIISYFTFOKXSEILVNVETGL-----NFTCGSSNVSVTT 596
 DB 599 FFLNSASVPAVYFRTYLSLWSFRYANFALLINQSTVYDGIACRANVTCPSREILLET 657

RESULT 12

WHITE DROME

ID WHITE DROME STANDARD; PRT; 687 AA.

AC P10090; Q9V3A2; Q9XY33;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1991 (Rel. 20, Last annotation update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE White protein.

GN W OR EG:BACN33B1.1 OR CG2759.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Head;

RX MEDLINE=90221897; PubMed=2109311;

RA Pepling M., Mount S.M.;

RT Sequence of a cDNA from the Drosophila melanogaster white gene.;

RL Nucleic Acids Res. 18:1633-1633(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=85134655; PubMed=6084717;

RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;

RT DNA sequence of the white locus of Drosophila melanogaster.;

RL J. Mol. Biol. 180:437-455(1984).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=21100348; PubMed=1156992;

RA Lukatsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niva S., Yamamoto D.;

RT "Dual-tagging gene trap of novel genes in Drosophila melanogaster.;"

RL Genetics 157:727-742(2001).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amani M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Marklow G., Malsbenden N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.;"

RL Science 287:2185-2195(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN-Oregon-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,

RA Barrall B.G., Ferraz C., Vidal S., Brun C., Demallies J., Borkova D.,

RA Dreano S., Gloux S., Lelaure V., Motter S., Galibert F., Borkova D.,

RA Minana B., Katatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,

RA Modolell J., Peter A., Schoettler P., Werner M., Mounikioti F.,

RA Belnert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Darlamtsou A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,

RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of D.

RL Science 287:2220-2222(2000).

RN [6]

RP SEQUENCE OF 224-331 FROM N.A.

RX MEDLINE=89339145; PubMed=2503416;

RA Tearle R.G., Belote J.M., McKewen M., Baker B.S., Howells A.J.;

RT "Cloning and characterization of the scarlet gene of Drosophila

RL melanogaster.;"

RL Genetics 122:595-606(1989).

- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY

FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS

RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE

TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF

TRYPTOPHAN.

- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.

- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

- THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

EMBL: X51749; CAA36038.1; -

EMBL: X02974; CAA26716.1; -

EMBL: AB028139; BAA78210.1; -

EMBL: AE003425; AAF45826.1; -

EMBL: AL133506; CAB65847.1; -

EMBL: X76202; CAA53795.1; -

PIR: S07263; FYFW.

PIR: S08635; S08635.

Flybase: FBgn0003996; w.

InterPro: IPR003593; AAA_Arpase.

InterPro: IPR003439; ABC_transporter.

Query Match:	18.1%:	Score 602.5:	DB 1:	Length 687:
Best Local Similarity	28.8%:	Pred. No. 1.5e-34:		
Matches 180:	Conservative 131:	Mismatches 220:	Indels 95:	Gaps 19

```

0Y      66  ROLKDLVSAYESQIMOLIGSSGCKFTLLIDAMGR--LGAQFLEGEVYNGALRRE 123
Db      110 KHLKNCGVAVPPELLAVMSSGSGAKFTLLNALFRSPDQIGVPSGRLNLGDPVAK 169
0Y      124 QFOQCFSTVLOSPTLLSLVYRETHYALLAIRNGNSGSPK---KVEAVAEISLSHV 180
Db      170 EMQARCAVVOODDLEFISLTAREHLEFQAM--VBRPHLTRYRQARARDOVIGELSLSC 227
0Y      181 ADRLIC-NYSLGIGSTGERRRVSTAOLLODPKWLFBEPPTGLDCMTANOIVLLVELA 239
Db      228 QHTTIGVPRKKGISGGERKRLAAASEALTBPPLICBEPSTGDLSTFAHSVVOYLKLS 287
0Y      240 RRNRIVLLTHOPRSELEFOLDRLTAISFGELIFCGTPEAMIDPDCGYRPPESNPD 299
Db      288 OKGTVLTLTHQPSSELEFELDKLLMAEGVAFGLTSEAVDFSTYGAOCPTYNPAD 347
0Y      300 FYMDLTSVDVQSKRELETSKRVOMIESAYKKAICHKTLNIEBKHLKTLPPVPFKTK 359
Db      348 FYQVAVLAV---VPGREIESRDIRIKICDNFAIS---KVARDMEQLATKNLE---KPL 396
0Y      360 DSP-----GVESKIGVLLRRTYRNRKRLAVITRLLONLIMGLEFL-LEFVLKRVSN 410
Db      397 EOPENGTYTATWEMQPRAYAMKRWLSLAKPELLKYNLIDTTWAILLIGLIFLQOULO 456
0Y      411 VLGAIQDRVGLYQVGAATPYTGMLANVNEPVLRAVSDQESODGLQOKOMOMALVALH 470
Db      457 V---GVANNIGALFLPLTNMFWQVAVPATINFTSELPTVMEARSRLKRCOTYFLGKITA 513
0Y      471 VLPPSVATMIFSSVCWTGLGHEVARAFGTFSSALLAPHLIGLEFLTVLLGIYONPVIV 530
Db      514 ELPELEFVLPVPEFTAIAPMIGLRAGVYHF-----FNCIALVTVL--ANVS 556
0Y      531 NSVALALISAG-----VLVSGFLRNQIEMPIPKIISYTFPOKCEI 574
Db      557 TSGCYLISCASSSTSMALSVGPPIITPFLLEGGFLLSGSPVYIKWLSYLSWRYAEG 616
0Y      575 LVVNEFYGL--NFTCGSSNVSVTTPNMCAPFTQIOPIEKTCP--GATSRETNLILYS 629
Db      617 LLINOMADVPEGEISCTSSNT-----TCSGSKVILETTLNFS--A 655
0Y      630 FIP-----ALYIIGVYKIRDHLSR 651
Db      656 DLPLDYVGLAIL-IVSFRVLAIALR 680

```

ID ABG1_HUMAN STANDARD; Part: 678 AA.
 AC P45844; Q9BXK0; Q9BXK1; Q9BXK2; Q9BXK3; Q9BXK4; Q9BXK5; Q9BXK6; Q9BXK7;
 AC Q9BXK8; Q9BXK9;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABC8 OR WH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Retina;
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Lalliot M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the Drosophila white
 RT gene and mapping to chromosome 21q22.3";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Bleichmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehman R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordiolek G., Hornischer K., Brandt P.,
 RA Schafte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramseyer S., Borzym K., Klages S., Hennig S., Rieseemann L., Dagand E.,
 RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Leharz H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Talor I., Korostishevsky M.,
 RA Wattenhofer M., Galipponi M., Barris C., Rossier C., Shibuya K.,
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness
 RT DENB10 locus using 34 novel microsatellite markers: genomic
 RT structure, and exclusion of six known genes in the region.";
 RL Genomics 66:22-29(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=21192304; PubMed=11279031;
 RA Poroch-Oezuerumez M., Langmann T., Helmerl S., Borsukova H.,
 RA Kaminski W.E., Drobniak W., Honer C., Schumacher C., Schmitz G.,
 RA "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
 RA of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
 RA expression and a modulator of cellular lipid efflux."
 RL J. Biol. Chem. 276:12427-12433(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3, 4, 5, 6 AND 7).
 RC MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene";
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
 RN [6]
 RP SEQUENCE OF 33-678 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Arclnlegas S., Son D., Wu R.;
 RT "Isolation and characterization of a mammalian homolog of the

Db 248 DSACFOVSLAKGLAOGRSIICTIHOPSAKLEFELFDQVLVSOGOCVYRGKVCNLVY 307
 QY 284 FNDGCPPEHSPNPFDFYNDLTSDYDTSKEREIETSKRVOMIESAYKSA-----I 334
 Db 308 LRDGLNCPYTHNRPADFEVAVASGEYGOQNSRLVRAVEGCDSDHKDLGDAEVPFL 367
 QY 335 CHKTUKNIEMKHKLTLPVDPKTPKDSPGV-----FSKGLVLLRYTVRLVNRKL 384
 Db 368 WHPRESEVKQYRKRLKGL-----RKDSSMEGCHSFSAKCTGOCILFKRFLSLIMDSV 421
 QY 385 AVITRLLONLMGFLFELFVLRVRSNVLKGAIDRVGILLYFVGATPTPYGLMNAVNLFPV 444
 Db 422 LTHLRITSHIGLIGLILLYGIGNEAKK--VLSNSGFLFESMFLMEALMPVLTPEPL 479
 QY 445 LRAVSDQSDGLTOKNOMLAVLALHVLPEFSVAPMIFSSCYMTGLHPEVARGYSA 504
 Db 480 EMGVFLREHLNMYSLAKAYLAKTMADVPFQIMPPVACSTIVYMTSOPSDAVRFLFA 539
 QY 505 ALLAPHLIGFELTVLLGIQVONPNIVNSVALLSIAGLVSGFLRNIOEMPIPEKISY 564
 Db 540 LGTWTSLVAOSIAGL-LIGAASTSLQVAFVGPVTAIPVLLFSGFVSFDITPYLQNMYS 598
 QY 565 FTPOKYSGLIYVNEFGIANTGSSVSVTTNPMCAFTGCIQFIETKCGATSRFTMNF 624
 Db 599 ISYVRGFEVILS-ITGLD---REDLHCDIDETCFH-OKSEALIRELDVENAKLYLDF 652
 QY 625 LILYSFIPALVILGIYV--FKIR 645
 Db 653 IVLGIFFISRLIAVFLRKIR 675

RESULT 14
 ID WHIT_CERCA STANDARD; PRT; 679 AA.
 AC 017320:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Tephritidae; Tephritidae; Ceratitis.
 OC NCBI_Taxid=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9613276; PubMed-8533095;
 RA Zyldebel L.J., Saccone G., Zacharopoulos A., Besansky N.J.,
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
 RT "The white gene of *Ceratitis capitata*: a phenotypic marker for
 RT germ-line transformation".
 RL Science 270:2005-2007(1995).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X89933; CAA61998.1;
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR005284; pigment_permease.
 DR Pfam: PF000005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transport; 1.
 DR TIGRFAMs: TIGR00955; 3a01204; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 121 128 ATP (BY SIMILARITY).
 FT TRANSMEM 427 445 POTENTIAL.
 FT TRANSMEM 457 477 POTENTIAL.
 FT TRANSMEM 507 525 POTENTIAL.
 FT TRANSMEM 534 555 POTENTIAL.
 FT TRANSMEM 568 586 POTENTIAL.
 FT TRANSMEM 651 670 POTENTIAL.
 FT CARBOHYD 628 628 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 679 AA; 75145 MW; 3P9C8C78A835C40C CMC64;
 Query Match 17.8%; Score 591; DB 1; Length 679;
 Best Local Similarity 28.4%; Pred. No. 9.4e-34;
 Matches 176; Conservative 125; Mismatches 231; Indels 88; Gaps 18;
 QY 66 RQIKDYSLVYESGOIICIGSSGSKTTLIDAMSGRLGRAGTFLGEV--YVNGRALRE 123
 Db 101 KHLKNDGSAVPEBELLAWMSSGAGKTTLLNABAFSSKGVQISPTIRMLNHPVDAR 160
 QY 124 QPDCFSYVLOSDFLLSLVRETLYHTALAIRNGPSFOK--KYEAVAEISLSTVA 181
 Db 161 EMQARCAVYQDDLEFISLAREHLIQAMVRMR-HHTQKQKQVARDQVQIDSLGRQ 219
 QY 182 DRLIG-NYSLGISTGERRRYSIAQLQDPRKMLFDEPTTGDCMTANQIVLVLELAR 240
 Db 220 NTLIGVGRVYGLSGGRKRLAFSEALTPPLLCDEPTSGIDSFMAHVVQYIKLSQ 279
 QY 241 RNRIVLTIHOPRSELFDLFDKTAIISFGLICGCTPAEMDFPNDGCPPEHSPNPF 300
 Db 280 KGTIVLTIHOPSELELFDKTLMEGRVAFGTGEAVDFSYIGATCTPYTPADF 339
 QY 301 YMDLTSDYDTSKEREIETSKRVOMIESAYKSAICHTKLNIEEMKHKLTLPVFPKTD 360
 Db 340 YVQVLAV---VPGREVSRRVAKICQNFVAGKYSREMEQNFQ---KLYKSGFKED 391
 QY 361 -----SPGVSKILVLLRVTNRLVNRKLAIVITRLLQNTIMGLFL-LFVLRVRSNVLK 413
 Db 392 ENETYKASWFMQFRAVLMRSWLSVLEKPELLVKKVRLQTTMVAVLIGLIFGQQLQV-- 449
 QY 414 GATDDRGLIXOPFAGAPPYGMLNAVLFVLRVSVQESQDGLYQKRMALAVLPL 473
 Db 450 -GVANINGALFELFTNMTFQNSPATTIVFTELPVFKRERSRYRCDYFLGKTIALP 508
 QY 474 ESVVATMIFSSVCWTGLHPEVARFEGYFAALLAPHLIGFELTVLGIQVONPNIVNSV 533
 Db 509 LFLVVPFLFAINAVPLGLRPGVDH-----FRLALVTLV--ANVSTSF 551
 QY 534 VALIS-----IAGVLVSGFLRNIOEMPIPEKISYFTPOKYSGLIYV 577
 Db 552 GYLISCASCSSTMAISVGPVLIIPFLFGLGFFLNSGVSVPYFKWLSTLSMFRYANEGLLI 611
 QY 578 NEFYGL--NFTGSSVSVTTNPMCAFTGCIQFIETC--GATSEFTANFL---ILYS 629
 Db 612 NQWADWAPGEITGLTSLT-----TCSSSGEVILETINFASDLPFD. 652
 QY 630 FIP-ALVILGIYVFKIRDL 648
 Db 653 FIGALLIIVG--FRISAVI 669
 RESULT 15
 ID ABG4_HUMAN STANDARD; PRT; 646 AA.
 AC 09H172:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 4.
 GN ABG4 OR WHITE2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21518231; PubMed-11606068;
RA Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,
RA Cullen P., Asmann G.;
RT "The human ABCG4 gene is regulated by oxysterols and retinoids in
RT monocyte-derived macrophages";
RL Biochem. Biophys. Res. Commun. 288:483-488(2001);
RN [2]
RP SEQUENCE OF 20-646 FROM N.A.
RC TISSUE-Dorsal root ganglion;
RA Oldfield S., Lowry C.A., Lighman S.L.;
RT "Cloning and expression of a novel mammalian white family
RT ABC transporter: WHITE2";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in macrophage lipid homeostasis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ300465; CAC17140.1; -
CC Genew; HGNC:13884; ABCG4.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Glycoprotein; Transmembrane; Transport.
KW DOMAIN 1 393
FT TRANSMEM 394 414
FT DOMAIN 415 425
FT TRANSMEM 426 446
FT DOMAIN 447 472
FT TRANSMEM 473 493
FT DOMAIN 494 503
FT TRANSMEM 504 524
FT DOMAIN 525 532
FT TRANSMEM 533 553
FT DOMAIN 554 617
FT TRANSMEM 618 638
FT DOMAIN 639 646
FT NP_BIND 102 109
FT CARBOHYD 422 422
SQ 646 AA; 71895 MW; 9CCEC6E150772611 CRC64; (POTENTIAL).

Query Match 17.4%; Score 578.5; DB 1; Length 646;
Best Local Similarity 27.1%; Pred. No. 6; 6e-33;
Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;

OY 33 PEPHSIGILHASTYSYSHRRPMMWDITSCROQWTRQILKDVSLYVESGQIMCITLSSGSGK 92
DB 54 PARSAVDIEFVELSTSVREGPCW-----RKRGYKTLKCLSGKFCRRELIGIMGPSGAK 108
OY 93 TTLIDAMSGRLGACFTLGEVYVNGRALREPODFSVYLSQDITLSSLYVRETLHVTYA 152
DB 109 STEFNILAGY--RESGMMGQILYVNGRPRELRTFRKMSCTIMQDDMLPHLYLLEAMVSA 166
OY 153 LLAIRGNFGSFOKVEAV-----MAELISHVADRLIGNSLIGISTGERRVSIQAQ 206
DB 167 NKLISEKO-----EVKKEIVTEITLALGIMSCSHRTAL-----LSGGQRRLAIALE 214
OY 207 LLQDPKVMFLDEPTSLGDCMTANQIVLVLLVELARNRIVVLTIHQPRSELPLEPKIAL 266

DB 215 LVNPPVMEFDEPTSGLDNSASCFOVVSIMKSLAOGGRITICTIHQPSAKLFEFEDLYLI 274
OY 267 SFCELIIFCGTAPALMEFNDGYPCEHNSPPDFVYDITLSDQSERELIETSKRYOMIE 326
DB 275 SQGQCIFKGVYVTLIYIKLGILGACPTYNHPADFIIEVASG-----EGDLPMPLE 325
OY 327 SAYKSAICHTKILKNIERMKHLKTLPMVPEFKTQSD-----GVFSKIGVLLARVTRN 378
DB 326 RAV-QNGLCMAEKKSSPEKNEVPAPCPPOPEVDIEHSTFATSTLQFCILFKRTIFS 384
OY 379 LVNKLAVITRLQNLIMGLFELFVLRBSNVLKAIQDRGLLYQVGAAPPYGMLAN 438
DB 385 ILNDTVLTHLRFMSHVIVGLIGLILYHIGDDASK--VFNNMGCLFFSMILFALMLPT 442
OY 439 VNLFPVLRVSDQSDGLYQKQMMALAYALHLPSVYATMFSSVCWTIGLHEVAR 498
DB 443 VLFPLDMAVPMFELHNTWYSLKAYILATMDVPPQVVCISIVMTMGQPAETSR 502
OY 499 FGYSALLAPHLIGELTLVLLGIVQNPNIYNSVALLSIGVLGSGFLRNIOEMP 558
DB 503 FILFSALATATALVAQSLG-LIGAASNSIQVATFVGPYTAIPVLFSGFFVSFKTIPT 561
OY 559 FKLIISFTYQKTCSEILVNEFYGL--NFGSSNVSVYTNPMCAFTQGIQFIERTCPG 615
DB 562 LQWSSYLSYVRGFEVGIILT-IYGMERGDLTC-----LEERCPRFP-QSILRALDY 611
OY 616 AYSRFTMNLILYSFIPALVYIIVYFKIR 645
DB 612 EDKLYMDPLVJGIFPLALRLAYLVLR 641

Search completed: June 11, 2003, 09:12:37
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:11:47 ; Search time 18 Seconds

(without alignments)
1064.129 Million cell updates/sec

Title: US-09-989-981A-6

Perfect score: 3326

Sequence: 1

BL0SUM62

Scoring table: Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, AA:

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/CTUS.COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/Dackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682.5	20.5	655	4	US-09-245-808-1
2	244	7.3	248	4	US-09-134-001C-3731
3	244	7.3	1280	2	US-08-752-447-2
4	244	7.3	1280	4	US-09-316-167-2
5	240	7.2	1279	2	US-08-784-649A-2
6	240	7.2	1280	2	US-08-583-276-19
7	240	7.2	1280	6	5206352-4
8	239.5	7.2	1684	3	US-08-665-259-25
9	239.5	7.2	1684	3	US-08-762-500-25
10	239.5	7.2	1704	3	US-08-762-500-75
11	230.5	6.9	242	4	US-09-134-001C-3832
12	229	6.9	460	4	US-09-134-001C-3369
13	229	6.9	1307	1	US-08-395-246C-2
14	220.5	6.6	273	4	US-09-134-001C-3641
15	220	6.6	1457	3	US-08-665-259-27
16	220	6.6	1457	3	US-08-762-500-27
17	218	6.6	1275	3	US-09-120-513-2
18	218	6.6	1275	4	US-09-450-105-2
19	213	6.4	228	4	US-09-134-001C-3373
20	211	6.3	270	4	US-09-134-001C-5561
21	210	6.3	382	4	US-09-134-001C-3439
22	209.5	6.3	244	4	US-08-919-573-2
23	209.5	6.3	244	4	US-08-919-573-4
24	209.5	6.3	341	4	US-09-134-001C-4042
25	208.5	6.3	202	4	US-08-858-207A-332
26	208.5	6.3	229	4	US-09-134-001C-4113
27	207.5	6.2	246	4	US-09-627-376-15

28	206.5	6.2	233	4	US-08-993-825-2	Sequence 2, Appl1
29	206.5	6.2	1476	4	US-09-256-703-2	Sequence 2, Appl1
30	206.5	6.2	1480	1	US-07-637-621-2	Sequence 2, Appl1
31	206.5	6.2	1480	2	US-08-951-912-2	Sequence 2, Appl1
32	206.5	6.2	1480	4	US-08-681-838A-2	Sequence 2, Appl1
33	206.5	6.2	1480	4	US-08-681-838A-3	Sequence 3, Appl1
34	206.5	6.2	1480	4	US-09-174-077-2	Sequence 2, Appl1
35	206.5	6.2	1480	6	5240846-5	Patent No. 5240846
36	206	6.2	406	4	US-09-134-001C-3570	Sequence 3570, Ap
37	206	6.2	1479	2	US-08-951-912-4	Sequence 4, Appl1
38	206	6.2	1479	4	US-09-174-077-4	Sequence 4, Appl1
39	205.5	6.2	1480	4	US-09-425-453A-6	Sequence 6, Appl1
40	205.5	6.2	1480	4	US-09-425-453A-12	Sequence 12, Appl1
41	205.5	6.2	1480	4	US-09-425-453A-20	Sequence 20, Appl1
42	205	6.2	225	4	US-09-134-001C-3215	Sequence 3215, Ap
43	204.5	6.1	1308	2	US-08-996-644-2	Sequence 2, Appl1
44	204.5	6.1	1308	3	US-09-352-552-2	Sequence 2, Appl1
45	204.5	6.1	1480	1	US-08-136-742A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1	US-09-245-808-1	Sequence 1, Application US/09245808
Patent No. 6313277		
GENERAL INFORMATION:		
APPLICANT: Doyle, L. Austin		
APPLICANT: Abuzzo, Lynne V.		
APPLICANT: Ross, Douglas D.		
TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which		
FILE REFERENCE: Ross UMD conversion		
CURRENT APPLICATION NUMBER: US/09/245,808		
CURRENT FILING DATE: 1999-02-05		
EARLIER APPLICATION NUMBER: 60/073763		
NUMBER OF SEQ ID NOS: 7		
SOFTWARE: Patentln Ver. 2.0		
SEQ ID NO 1		
LENGTH: 655		
TYPE: PRT		
ORGANISM: Human MCF-7/Adrtp cells		
US-09-245-808-1		
Query Match	20.5%; Score 682.5; DB 4; Length 655;	
Best local similarity	29.2%; Pred. No. 1.9e-64;	
Matches	182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;	
OY	21 SSSLEGAPATAP---EPHSIGILHASYSVSHRPMDITSCROOITKIDSLYE 77	
DB	13 SGGNINFPATASNDLKAFTGCAVLSFNHCYRWKLSGFLRCRPFVEKILSNIGIMK 72	
OY	78 SQIHCIIASSSGGTTLLDANSGLRGAGTFLGEYVNGALRREQFDCSYVQSDT 137	
DB	73 PG-LNAILLPGTGGGSSLLDYLAAKRDSCG-LSGCVLLING-APRANFCNGSYVQDDV 129	
OY	138 LLSLTVRETLTYLLAIRGNPG-SPOKYEAVMAELSLSHVADRILGNISLGISTG 196	
DB	130 VAGTIVNENLOFSALRLATMTMHENKERNIRVIOELGIDKVDKSGQFINGVSG 189	
OY	197 ERRRVISAOLLDQPKVMFLDEPTGLDPMANOIVLVLLVLRNRRIYVLIHQPSRL 256	
DB	190 EKKRISIGMEITDPSIIFLDEPTGLDOSTANAVLLIKRSGKGRITISIHPRKSI 249	
OY	257 FQLEPKIILISFGEILFPGTPEMLDFPNDGCGPCPEHSNPFDFYMDLTSDVTO---SK 312	
DB	250 FKLFDSTLTLASGRIMFGPAQOEALGYFESAGYHCEAVNNPADFLDIINGDSTAVALNR 309	
OY	313 EEE-----IESKR-----VQIESAYKKSACHT-----LKNIEKKHKTLPMPWF 356	
DB	310 EEDFKATITIEPSKODKPLIEKLAETIYVNSSYFKETKALQLSGEKKRKITVKEISY 369	

QY 357 KTQSPGVSKGLVLRVTVNLVNNKLAIVTLLQNLGLFT--LEFVLRVSNVLKG 414
DB 370 TT-----SFCQHLRWKSKRSFKNLGNPQASIAQIIVTVVLGLVIGAIYEGKJNDST----- 421
QY 415 AIQDVGGLIYQVQATPTTGMLNANVLEFVLRAVSDOSQGLQKQKQMLAYAL-HVLP 473
DB 422 GIQNDAGVLE-FLTTCQFSSVSAVLEFVEKLEFHEHYISGYRVSSYFLGKLKLDLPL 480
QY 474 FSVVATNIFSSVCYWTGLLHPEVAFRGFYSAALLAPHLIGELFLVLLGIQVQNPVANSV 533
DB 481 WMLPSTIIFCTIYVFMGLKRAKADAFEMFTLM---WVAVSASMALAIAGOSVSA 537
QY 534 VALLSIACV--LVGSGFLRNTQEMPIPKIISTYFFQKCEIIVVNEFYGLNFTCGSSN 591
DB 538 TLMATICEFVFMIFSGLLVNLTTIASMLSMLOYSIPRYGTALQHNELQNCQPG--- 594
QY 592 VSVTTPMCAFTQGIQFLEKTCPG 615
DB 595 LMATGNPCMYA-----TCTG 610

RESULT 2

US-09-134-001C-3731
Sequence 3731, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134, 001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3731
LENGTH: 248
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3731

Query Match 7.3%; Score 244; DB 4; Length 248;
Best Local Similarity 29.3%; Pred. No. 8,4e-18;
Matches 70; Conservative 53; Mismatches 92; Indels 24; Gaps 9;

QY 67 QILKDVSLVYSGQIMCIISSGSGKTTLLDMSG-RLGRAGTFLGEVYVNGRALLRE-- 123
DB 20 EVIKGIDLKINQGVVLLIGRSGSKTLLLMINALEIPEGT---VYVGMTYNTKDK 75
QY 124 ----QFODCFSYVLOSITLSSLVRETLHTALLAIRGNPSGFQKKEVAVMELSLSH 179
DB 76 KSQLKTVQOQSGMYQNTNLPFKSALENV-MEGILTVKMKKATANEAMNLAKVGIVH 134
QY 180 VADRLIGNYSLSIGISTERRRVSIAQLDPPKYMLEDDEPTGLDCTANQIVLVLELA 239
DB 135 VKQD--RPHALSG---GQOQRAVALARALANPKMLDEPTSADEPVLNDYLVKIELA 189
QY 240 RRRNIVYLTIHQPSSELPQFLDKALISFGEILFCGPAEMLEDFNDGCGPCPSHNP 298
DB 190 DEGMTAVIVTHEMFAF-EVSNQIAFTHEGVIAEGQTPD--DIFN--HPKTELORF 241

RESULT 3

US-08-752-447-2
Sequence 2, Application US/08752447
Patent No. 5994088
GENERAL INFORMATION:
APPLICANT: Mechettner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and

TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: McDowell Boehnen Hubert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,447
FILING DATE: 15-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5994088nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-9808
TELEFAX: 312-913-9808
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-447-2

Query Match 7.3%; Score 244; DB 2; Length 1280;
Best Local Similarity 20.7%; Pred. No. 1.5e-16;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

QY 41 LHASYSYSHRYRPMWMDITSCNQWTRQILKDVSLVYSGQIMCIISSGSGKTTLLDMS 100
DB 397 VHFSPSRKEVK-----ILKGLNLTQSGGVTVLVGNSGCGSKSTVQIOL 441
QY 101 GRIGRAGTFLGEVYVNGRALLR--REQFODCFSYVLOSITLSSLVRETLHTYLLAIR 153
DB 442 -RL--YDPTBGMVSDQDQIRITINVRFLRELIGVSGEPVLFATTINENRY----- 490
QY 159 GNPSPQKRYAVNAE-----LSLHVADRLIGNYSLGISTGERRRVSIAAQLQDP 211
DB 491 GRENVTDELEKAVKANAYDFIMKLPKEDTLVGERG-AQLSGGQKORIAARALVRNP 549
QY 212 KVMLEDEPTGLDCMTANOIVLVLELARNRIVVLTIHQPSSELPQFLDKALISLSEL 271
DB 550 KIILDEATSLMD-TESEAVVQVADKARKGRTTIVLAH--RFAVFNADVIAGFDGVI 606
QY 272 IFCGTPAEMLDFFNDGCGPCPSHNPDPFYMDLTSVDQSKEREIETSK----- 320
DB 607 VERGNHDELM-----KEKGIFKLYTMQTAGNEVLENAADESKSEIDALE 652
QY 321 -----RVQMISATKSKATICKTKANIERMKHLLTPVVF-----KTQSPG 363
DB 653 MSSNDSRSLIRKSTRSVRSVGOQHRKLTREALD--ESIPVSFWRIMKMLNLTMPY 710
QY 364 -----VFSK-LGVLR-----RVTRN-----LYRNKLAVITR 389
DB 711 FVVGVCFAIINGGLQPAFAIIFSKIIIGVPRIDDEPKRONSLSLFLALGIITSTTP 770
QY 390 LQNLING-----LFLFVYLRVRSNV-----LKGAIQ 417
DB 771 FLOGFTFGKAGEILTKRLRYVFRSMRGVSWFHDKNKTGALTRLANDAGVYGAIG 830
QY 418 DRV-----GLLYQFVQATPTTGMLNANVLEFVL-----RAYSQDQ 452
DB 831 SRLAVITONTANIGTGIITISFYIGWQTLTLLAI--VPITAIAGVEMKFAQALNDK 888

QY 453 SODGL-----YOKWOMLAVAHY-----LPSVATM 480
 DB 889 ELEGAGKATEAENFRVVSUJOQFHEHYAOSIAQVYRNSLRKAHIFGITSFTQAM 948
 QY 481 IFSSV--CYWTLGLHPEVARFGYSAALLAPHLIGER--LTLVLGIYONRNIVNSV--- 533
 DB 949 MYFSYACGF-----RFG---ATLVAKRLMSFEDVLVFSAVFGAMAVGOVSSF 994
 QY 534 -----VALLSIAGLVGSGFLRNIOEMPIPKIISYTFQKYSFELVYNEFGYGLNFTCG 588
 DB 995 APDYAKAKISAHHIM-----IIEKPTL-----IDSYTEGLMPNTLEG-NTVTEG 1038
 QY 589 SSNSVTTNPMCAFTGCI 606
 DB 1039 EVENNYPTRPDIPLYQGL 1056

RESULT 4

US-09-316-167-2
 : Sequence 2, Application US/09316167
 : Patent No. 6365357
 : GENERAL INFORMATION:
 : APPLICANT: Mechetter, Eugene
 : APPLICANT: Ronlison, Igor B
 : TITLE OF INVENTION: Methods and Reagents for Preparing and
 : NUMBER OF SEQUENCES: 2
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
 : STREET: 300 South Wacker Drive, Seventh Floor
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: USA
 : ZIP: 60606
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentia Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/316,167
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/752,447
 : FILING DATE: 15-NOV-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: No. 6365357nan, Kevin E
 : REGISTRATION NUMBER: 35,303
 : REFERENCE/DOCKET NUMBER: 95,1121
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 312-913-0001
 : TELEFAX: 312-913-9808
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1280 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-09-316-167-2

Query Match 7.3%; Score 244; DB 4; Length 1280;
 Best local Similarity 20.7%; Pred. No. 1.5e-16;
 Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;
 QY 41 LHASYSVSHRVRPMWDITSCROQWTRILKDVSLYVESGQIMCILSSGSGKTTLDAMS 100
 DB 397 VHSYSPSRKEVK-----ILKGLMLKQSGFTVALVNSGSGKSTVQIOMQ 441
 QY 101 GRGRAGTFLGEYTVNGRALR--REOFQDCFSYVLSDVTLSSLVARETLHYTALAIRR 158
 DB 442 -RL--YDPTGEMVSDGDIRITINVRFLREITIGVVSQEPVLFATTAENIRY----- 490

QY 159 GNPGSFQKVEAVNAE-----LSLSVADRILGNSTSGISGERRRSIAAQLDDP 211
 DB 491 GRENVTMEDEKAEVANAAYPIIMKLPKFTPLVGEERG-AOLSGQKORIAIAALVNP 549
 QY 212 KVMLEDEPTTGIDCTANQIVLVLELARNRNIVLTHOPRSEIFQFEDKIALISFGL 271
 DB 550 KILLDEATSLD--TESAVVQVALDKARKRTTIVIAH--RFTVRADYIAFGDGYI 606
 QY 272 IFCTPAPMLDFNDCCGIPCEPHSNPEFYDLTSVDTQSKERELETISK----- 320
 DB 607 VEKGNHDELM-----KENGIFYELVTMOTAGNEVELENAADSKSEIDALE 652
 QY 321 -----RYOMIESAYKSAICHTLNIERMKHLKLPVPP-----KTKSPG 363
 DB 653 MSSDSNSSLIRKSTRSRVRSQAOHRKISTKEALD--ESIPVSEWRIRKMLNLTMPY 710
 QY 364 -----VESK-LGVLR-----RVTRN-----LVNKLAVTR 389
 DB 711 FVGVFECAIINGGLQPAFAIFSKIGVFTRIDDEPTRYKNSNLFSLFLALGIISFTTF 770
 QY 390 LLQNLIMG-----LELIFVLRVNSV-----LKGAIQ 417
 DB 771 FLOGFTFGKAGEIILTKRLRYVWFRSMLEQDVSFHPDKNTGALTTRLANDAAYKGAIG 830
 QY 418 DRV-----GLIYQVGAIPYTGMLNAYNLPVL-----RAVSDQ 452
 DB 831 SRLVITQONIANLGTGIIISFIYGMQLTLILLAT--VPIIAIAGVEMKMPACQALDKK 888
 QY 453 SODGL-----YOKWOMLAVAHY-----LPSVATM 480
 DB 889 ELEGAGKATEAENFRVVSUJOQFHEHYAOSIAQVYRNSLRKAHIFGITSFTQAM 948
 QY 481 IFSSV--CYWTLGLHPEVARFGYSAALLAPHLIGER--LTLVLGIYONRNIVNSV--- 533
 DB 949 MYFSYACGF-----RFG---ATLVAKRLMSFEDVLVFSAVFGAMAVGOVSSF 994
 QY 534 -----VALLSIAGLVGSGFLRNIOEMPIPKIISYTFQKYSFELVYNEFGYGLNFTCG 588
 DB 995 APDYAKAKISAHHIM-----IIEKPTL-----IDSYTEGLMPNTLEG-NTVTEG 1038
 QY 589 SSNSVTTNPMCAFTGCI 606
 DB 1039 EVENNYPTRPDIPLYQGL 1056

RESULT 5

US-08-784-649A-2
 : Sequence 2, Application US/08784649A
 : Patent No. 5830697
 : GENERAL INFORMATION:
 : APPLICANT: SIKIC, Branislav I
 : APPLICANT: Chen, Gang
 : TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
 : NUMBER OF SEQUENCES: 5
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Fish & Richardson
 : STREET: 2200 Sand Hill Road
 : CITY: Menlo Park
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94025
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentia Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/784,649A
 : FILING DATE:
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Sherwood, Pamela J

REGISTRATION NUMBER: Reg. No. 5830697 36, 677
 REFERENCE/DOCKET NUMBER: 06037/007001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-322-5070
 TELEFAX: 415-854-0875
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1279 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-784-649A-2

Query Match 7.2%; Score 240; DB 2; Length 1279;
 Best Local Similarity 20.7%; Pred. No. 4e-16;

Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

QY 41 LHASYSVSHRVPWMDITSCROQWTRQILKDVLSYVSGQIMCLSSGSGKTTLDAMS 100
DB 396 VHSYSPSRKEVK-----ILKGLMLKVOGQTVALLVNSGCGKSTVQLMQ 440
QY 101 GRGRAGTFLGEYVYNGRALR--REOFDCFSYVLSQDILLSSLVRETLHYTALLAIR 158
DB 441 -RL--YDPRGAVSVGDQDIRTNVFLREIIGVVSQEPVLEATTIAENIRY----- 489
QY 159 GNGSFQKKEAVMAE-----LSLSHVAURLIGNSLIGSISTGERRRVSTAAQLLODP 211
DB 490 GRENVTMDIEKAVKEANAYDFIMKLPKHFDTLVGERG--AQLSGGQKQRIAIRALVRNP 548
QY 212 KYMLPEPTGLDCAFTANOIVLVLEARNRIVLTIRPSELQLPDKIALLSGEL 271
DB 549 KILLDEATSAID--TESAVVOYALDKARKGRTTIVIAH--RLSTVRNADVIAGFDGVI 605
QY 272 IFCGPAEMLDFEFCGYPCEPHSNPFDFMDLTSVDQSKEREIETS-----KRVQ 323
DB 606 VEKGNHDELM-----KEKGIYFKLVYVOTAGNEVELENADESKSEIDALE 651
QY 324 MIESAKKSAICKTKTKNTERMKHL-----KTLPMVPF-----KTKDSPG-- 363
DB 652 MSSNDSRSSLLIRKSTRRSVSGSQADRKISTKEALDESIPVSFVRIMKMLNTEMPYEV 711
QY 364 -----VESK-LGVLLR-----RYTRN-----LYRNKLAIVTRLL 391
DB 712 VGVFCALINGGLQAPAFALIFSKIGVFTNIDPETRKNSNLSFLALGIIISFTFPL 771
QY 392 QNLIMG-----LFLLEFVLRVSNV-----LKGAIDOR 419
DB 772 QGFTGKAGEILLKRLRWVFSMLRQDVSNFDDPKNTGALTTRLANDAAQVKAIGSR 831
QY 420 V-----GLLYQVGVATPYTGMNAVNLFVPL-----RAVSDQESQ 454
DB 832 LAVITONIANLGTGIIISFIYGNQLLLLLAI--VPIIAIAGVEKMKLSGQALKDKREL 889
QY 455 DGL-----YOKOMMLAYALNV-----LPSVVAITMF 482
DB 890 EAGKATATEAIEFTVVSLSLQDQKEHMYAQSLOVYRNSLSLKAHIFGTFSETOAMMY 949
QY 483 SSV--CYMTLGLHPEVARGFSALLAPHLIGEF--LTLVLGLIYQONNIVNSV----- 533
DB 950 FSYAGCF-----RRG-----AVLVAHKLSMSPEDVLVSAVYFGMAVGOVSSFP 995
QY 534 ---VALLSIAGVLVSQGLRNIOEMPIPKIISYTFQYKCSSEILVNEFYGLNFCGSS 590
DB 996 DYAKKIKISAHHIM-----IIEKTPPL-----IDSYSTEGILMPNLTLEG--NATFGEV 1039
QY 591 NVSYTNPACAFTOGI 606
DB 1040 VENVTPREDIPVLOGL 1055

```

RESULT 6
 US-08-583-276-19

Sequence 19, Application US/08583276
 Patent No. 5837536
 GENERAL INFORMATION:
 APPLICANT: McDonagh, Kevin T.
 APPLICANT: Mienhuis, Arthur
 APPLICANT: Tolstoshev, Paul
 TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
 TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
 TITLE OF INVENTION: SELECTION OF CELLS TRANSFECTED WITH SUCH GENES
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Carella, Byrne, Bain, Giffillan,
 ADDRESSEE: Cecchi & Stewart
 STREET: 6 Becker Farm Road
 CITY: Roseland
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch diskette
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: DM4.V2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/583,276
 FILING DATE: 05-JAN-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/332,444
 FILING DATE: 31-OCT-1994
 APPLICATION NUMBER: 07/887,712
 FILING DATE: 22-MAY-1992
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1280 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE:
 DESCRIPTION: protein
 US-08-583-276-19

Query Match 7.2%; Score 240; DB 2; Length 1280;
 Best Local Similarity 20.7%; Pred. No. 4e-16;
 Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

QY 41 LHASYSVSHRVPWMDITSCROQWTRQILKDVLSYVSGQIMCLSSGSGKTTLDAMS 100
DB 397 VHSYSPSRKEVK-----ILKGLMLKVOGQTVALLVNSGCGKSTVQLMQ 441
QY 101 GRGRAGTFLGEYVYNGRALR--REOFDCFSYVLSQDILLSSLVRETLHYTALLAIR 158
DB 442 -RL--YDPRGAVSVGDQDIRTNVFLREIIGVVSQEPVLEATTIAENIRY----- 489
QY 159 GNGSFQKKEAVMAE-----LSLSHVAURLIGNSLIGSISTGERRRVSTAAQLLODP 211
DB 491 GRQNTVMDIEKAVKEANAYDFIMKLPKHFDTLVGERG--AQLSGGQKQRIAIRALVRNP 549
QY 212 KYMLPEPTGLDCAFTANOIVLVLEARNRIVLTIRPSELQLPDKIALLSGEL 271
DB 550 KILLDEATSAID--TESAVVOYALDKARKGRTTIVIAH--RLSTVRNADVIAGFDGVI 606
QY 272 IFCGPAEMLDFEFCGYPCEPHSNPFDFMDLTSVDQSKEREIETS-----KRVQ 323
DB 607 VEKGNHDELM-----KEKGIYFKLVYVOTAGNEVELENADESKSEIDALE 652
QY 324 MIESAKKSAICKTKTKNTERMKHL-----KTLPMVPF-----KTKDSPG-- 363
DB 653 MSSNDSRSSLLIRKSTRRSVSGSQADRKISTKEALDESIPVSFVRIMKMLNTEMPYEV 712
QY 364 -----VESK-LGVLLR-----RYTRN-----LYRNKLAIVTRLL 391
DB 713 VGVFCALINGGLQAPAFALIFSKIGVFTNIDPETRKNSNLSFLALGIIISFTFPL 772

```

QY 392 QNLING-----LFLFEVLVRSNV-----LKGAIODR 419
 Db 773 OGFEFGAGAGELTFRRLRYMFRSMRLRODVSWFDDPKNTGALTTRLANDAQAQVGAIGSR 832
 QY 420 V-----GLIYQVGAATPYTGLNAVNLFPVL-----RAVSDQ 454
 Db 833 LAVITONIANIGTGIISFIYGMQLTLLAL--VPILAIAGVEMKLSQALKDKKEL 890
 QY 455 DGL-----YOKOMMLATYALH-----LPSVATWIF 482
 Db 891 EGAGKATEALEINERTVVSLOTQOKFEHMTAOSIQVRYNSLRKAHIFGITFSTQAMT 950
 QY 483 SSV--CYWTLGLHBEVARFGYFSALLAPHLIGEF--LTVLIGIYONPNIVNSV-----533
 Db 951 FSVAGCF-----RFG-----AYLVAKHMSFEDEVLLVFSAVFGAAGVSSFP 996
 QY 534 ---VALLISGLVSGFLRNIOEMPPIFKIISYTFQKYCSEILVNEFYGLNFTGSS 590
 Db 997 DYAKAKISAHIIM-----IIERTPL-----IDYSTBGLMNTLEG--NVTGGEV 1040
 QY 591 NVSVTTPMCAFTQGI 606
 Db 1041 VFNPTRPDIPVLQGL 1056

RESULT 7

5206352-4
 Patent No. 5206352

Applicant: Konlison, Igor B.; Pastan, Ira H.; Gottesman, Michael M.

Title of Invention: COMPOSITIONS FOR CLONES CONTAINING DNA SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS

Number of Sequences: 4

Current Application Data:

Application Number: US/07/622,836

Filing Date: 24-SEP-1990

Prior Application Data:

Application Number: 892,575

Filing Date: 01-AUG-1986

Application Number: 845,610

Filing Date: 28-MAR-1986

Seq ID No: 4

Length: 1280

5206352-4

Query Match 7.28: Score 240; DB 6; Length 1280;
 Best Local Similarity 20.7%; Pred. No. 4e-16;
 Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

QY 41 LHASYSVSHRVRRMMWITSCROQMTNIIKDVSLVYESGQIMCIISSGSGKTTLDAMS 100
 Db 397 VHSYSRKREVK-----ILKGLMKVQSGQVALVSGSGKSTTVQLMQ 441
 QY 101 GRLGRAGTFLGEVYVNGRALR--REOFODCFSYVLSDTLSSLTFRHTLYTALLAIR 158
 Db 442 -RL--YDPEGMVSVGODIRTNVFLREIIIGVSOEPLFTTAENIRY-----490
 QY 159 GNGSFOKKEAVMAE-----LSLSHADRLIGNVSLIGISITGERRRVSIAAQLLODP 211
 Db 491 GRENVTMDEIKAVKEANAYDFIMKLPHEFDLYGREG--AOLSGQKORIAARALVYRN 549
 QY 212 KVALPDEPTTGLDCMTANQIYVLLVELARNRIVYLTINPSELQOLPKIILSFGEL 271
 Db 550 KILLDEATSAID--TSEAVVOVALDKARKGRTTIVIAH--RLSTVRNADVIAGFDGVI 606
 QY 272 IFCGPAEMLDEFNDGYPCEPHSNPFDFYMDLTVDYOSKEEIEITS-----KRVQ 323
 Db 607 VEGKNDELH-----KKGIIYFKLYVTQTAGNEVELENAADESGSEIDALE 652
 QY 324 MIESAYKKAICHKTKIKNTERMKH-----KTLPMVPF-----KTKDSPG--363
 Db 653 MSSNDSRSSLIRKSTRSRSVRSQAODRKISTFEALDESIPVSEFRIMKLNLTENPIYV 712

QY 364 -----VFESK--LGVLLR-----RVTRN-----LVNKKIATVITRLL 391
 Db 713 VGFCAIITNGLOPAPAIITFSKITGVTRIDDPETRKONSLFSLFLAIGIISFIYFPL 772
 QY 392 QNLING-----LFLFEVLVRSNV-----LKGAIODR 419
 Db 773 OGFEFGAGAGELTFRRLRYMFRSMRLRODVSWFDDPKNTGALTTRLANDAQAQVGAIGSR 832
 QY 420 V-----GLIYQVGAATPYTGLNAVNLFPVL-----RAVSDQ 454
 Db 833 LAVITONIANIGTGIISFIYGMQLTLLAL--VPILAIAGVEMKLSQALKDKKEL 890
 QY 455 DGL-----YOKOMMLATYALH-----LPSVATWIF 482
 Db 891 EGAGKATEALEINERTVVSLOTQOKFEHMTAOSIQVRYNSLRKAHIFGITFSTQAMT 950
 QY 483 SSV--CYWTLGLHBEVARFGYFSALLAPHLIGEF--LTVLIGIYONPNIVNSV-----533
 Db 951 FSVAGCF-----RFG-----AYLVAKHMSFEDEVLLVFSAVFGAAGVSSFP 996
 QY 534 ---VALLISGLVSGFLRNIOEMPPIFKIISYTFQKYCSEILVNEFYGLNFTGSS 590
 Db 997 DYAKAKISAHIIM-----IIERTPL-----IDYSTBGLMNTLEG--NVTGGEV 1040
 QY 591 NVSVTTPMCAFTQGI 606
 Db 1041 VFNPTRPDIPVLQGL 1056

RESULT 8

US-08-665-259-25

Sequence 25, Application US/08665259

Patent No. 6028173

General Information:

Applicant: Landes, Gregory M.

Applicant: Burn, Timothy C.

Applicant: Connors, Timothy D.

Applicant: Dackowski, William R.

Applicant: Van Raay, Terence J.

Applicant: Kilinger, Katherine W.

Title of Invention: NOVEL HUMAN CHROMOSOME 16 GENES.

Number of Sequences: 73

Correspondence Address:

Addressee: GENZYME CORPORATION

Street: One Mountain Road

City: Framingham

State: Massachusetts

Country: United States of America

Zip: 01701

Computer Readable Form:

Medium Type: Floppy disk

Operating System: IBM PC compatible

Software: Patent in Release #1.0, Version #1.30

Current Application Data:

Application Number: US/08/665,259

Filing Date: 17-JUN-1996

Classification: 435

Attorney/Agent Information:

Name: Dugan, Deborah A.

Registration Number: 37,315

Reference/Docket Number: IGS-9.1

Telecommunication Information:

Telephone: (508) 872-8400

Telefax: (508) 872-5415

Information for Seq ID No: 25:

Sequence Characteristics:

Length: 1684 amino acids

Type: amino acid

Topology: linear

Molecule Type: protein

MOLECULE TYPE: protein
US-08-762-500-25

MOLECULE TYPE: protein
5-08-762-500-75

MOLECULE TYPE: protein
5-08-762-500-75

Query Match 7.2%; Score 239.5; DB 3; Length 1704;
 Best Local Similarity 32.5%; Pred. No. 7.5e-16;
 Matches 78; Conservative 43; Mismatches 100; Indels 19; Gaps 9;

QY 66 RQILKDVSLVESQIICILGSSGSGKTTLLDAMSGRLGRAGTFLEGVYNGRALRR- 123
 DB 546 RAAVRDNLNLYEQEIVLVGHNGAGTTLTSLMTGLFPPTS--GRAVIGYEILSDMV 602
 QY 124 QFODCFSTVLOSPTLLSLVRETLHTALLAIRGNPGSFKQKEAVMAELSHVADR 183
 DB 603 QIRSLGICFQHDILFNLVVAEHLTFYQOL--KGL-SHQKPEEVKOMLHTIGLEDK 657
 QY 184 LIGNYSLGISTGERRRVSIQAOLLODPKVMLEFDEPTGLDCTANOIVVLELARRN 243
 DB 658 --WNSRSLFLSGMRRLTSLGIALINASKVILDEPTSGMDAIRRAIMOLL-QROKSDR 714
 QY 244 IVVITIHQPSSEFLQLEDKTALISFGLIFCGTPAEMLDFENDCGYPC---PERSNPF 299
 DB 715 TIVLTTHF-MDEADLDLDRALIMAKGELCCSSLFLKOKYG-AGYHMTIVKEPHCPED 772

RESULT 11

US-09-134-001C-3832
 ; Sequence 3832, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3832
 ; LENGTH: 242
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3832

Query Match 6.9%; Score 230.5; DB 4; Length 242;
 Best Local Similarity 26.1%; Pred. No. 2.3e-16;
 Matches 64; Conservative 56; Mismatches 90; Indels 35; Gaps 7;

QY 67 QILKDVSLVESQIICILGSSGSGKTTLLDAMSGRLGRAGTFLEGVYNGRAL----- 120
 DB 17 EYLRDNLNLYEKEVVAIIQPSGSKTLLRCMN-----LLDVPSGKVIYFENEL 67
 QY 121 ----RROFODCFSTVLOSPTLLSLVRETLHTALLAIRGNPGSFKQKEAVMAEL 175
 DB 68 TQHNHLDNLROKRGWFOFNLFPHKVIENYMAVL-LHKOSKQLEKALYLEKV 126
 QY 176 SLSHVADRLIGNYSLGISTGERRRVSIQAOLLODPKVMLEFDEPTGLDCTANOIVV 235
 DB 127 GLDOKAD---SLP-NOLSGGOKORVAIALAMEPVLMLEFDEPTSLADEVVGDIKVM 181
 QY 236 VELARRNRIVVLTIHQPSSEFLQLEDKTALISFGLIFCGTPAEMLDFENDCGYPCPEHS 295
 DB 182 ROLANESMIVVITHE-MNFAKISDKYVFMADGVYVESGTPONI-----FENQHS 232
 QY 296 NPEDF 300
 DB 233 RTENF 237

RESULT 12

US-09-134-001C-3369
 ; Sequence 3369, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3369
 ; LENGTH: 460
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3369

Query Match 6.9%; Score 229; DB 4; Length 460;
 Best Local Similarity 25.2%; Pred. No. 1e-15;
 Matches 77; Conservative 66; Mismatches 121; Indels 42; Gaps 10;

QY 66 RQILKDVSLVESQIICILGSSGSGKTTLLDAMSGRLGRAGTFLEGVYNGRALRR- 122
 DB 57 KKAVDNISLDIQSEFLAFTIGTSGSKTTL-RNINMIETD--GQIMNGKDYRRMNP 113
 QY 123 EORODCFSTVLOSPTLLSLVRETLHTALLAIRGNPGSFKQKEAVMAELSHVADR 182
 DB 114 VELRRSIGYVIOQIGLMPHTIREN-----IVLVPKLLKSKERKDEKARELKIYVLE 168
 QY 183 RLIGNYSLGISTGERRRVSIQAOLLODPKVMLEFDEPTGLDCTANOIVVLELARRN 242
 DB 169 EYLDRLP-AELSGGQORIVYRALAEODIILDEPFGALDPTTRTLDLVKELQOKL 227
 QY 243 RIYVLTIHQPSSEFLQLEDKTALISFGLIFCGTPAEMLDFENDCGYPCPEHSNPFY 302
 DB 228 GKTFIVTQHMDEKIRADKICIMSKKVVQYOPDNL-----RIPAN-----DFVR 275
 QY 303 DLTSVDQSKERELETSKRQMIKSAIKCHTKL-----NIEPMKHLKTLPM 353
 DB 276 DFI-----GONRLIDRPNMKSVESAMIKFV---TVKADSLDVAVINMTRVYTIIV 326
 QY 354 VPFKTR 359
 DB 327 VNNQNK 332

RESULT 13

US-08-395-246C-2
 ; Sequence 2, Application US/08395246C
 ; Patent No. 5773214
 ; GENERAL INFORMATION:
 ; APPLICANT: Peery, Robert B.
 ; APPLICANT: Skatrud, Paul L.
 ; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
 ; TITLE OF INVENTION: ASPERGILLUS FLAVUS
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Eli Lilly and Company
 ; STREET: Lilly Corporate Center
 ; CITY: Indianapolis
 ; STATE: IN
 ; COUNTRY: USA
 ; ZIP: 46285
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/395,246C
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

```

NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35784
REFERENCE/DOCKET NUMBER: X9683
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
TELEFAX: 317-277-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-395-246C-2

Query Match
Best local Similarity 23.0%; Score 229; DB 1; Length 1307;
Matches 127; Conservative 110; Mismatches 188; Indels 126; Gaps 29;

OY 68 ILKDVSLVYESGQIMCILLSSGSGKTTLLDMSGRIGR-AGFFLEGVY-----VNGRALRR 122
DB 407 VAHIDSCYIPAGKTAFAVGPSSGSKSTIISLERFYDPAAGTIMGHDIQTLNLRMLQ 466
OY 123 EFOFDFSYVLQSDTLSSLYRETLHYTALLAIRGNPSFO--KKEAV--ME-----174
DB 467 Q-----MSLYSQEPRLFAI-TIAENIRY-GIGSRFEKSTYEIRKRVPAARMAAHDF 519
OY 175 -LSLSHVADRLIGNSLGISTGERRRVSIAQLQDPKVMLEFDEPTTGDCMTANOIV 233
DB 520 IMALPGDYPTNIESFSLG---GOKORIMARAIIMDPKILLDETSALD-TKSEKLVQ 575
OY 234 LLVELARRRRIYVLTIHOPRSELPQEDKIALISFGEILFGSTPAEMLDFFNDCCYPCPE 293
DB 576 AALDKASKGRTTIVIAH--RLSTIQKANYIVIANQOIEQGPHEILMD-----622
OY 294 HSNPDFYMDLTSVTOSEKEREIETSRQVMTESAVKKSALCHKTLNTERKHLKTLPM 353
DB 623 ---RGIYCDMEAH-EIKKRSKRSKRSQI-----LTNISP---KHNP 661
OY 354 VPFKTKDSFG-----VF-----SKLGYLLRRVTRNLYRNKLAIVTRLLQMLINGLFTL 401
DB 662 TFFFKDYPGDESDIYSILSDASDIGHTEKQRPVSRMSLSHLMQPVKEAVSFWTL 721
OY 402 -----FFVLRVSRNVKLGATODRGILXOPGATPYTGMLNANVLEFVLRV 448
DB 722 FRTLASFRNPEPFLILGLCASISLAGIOPSQAVLEAFAVSTLSLPLE---YFKLR--775
OY 449 SDQESQDGLYQKQMLAYALHVPFVAVTMI-FSS--VCY-----WTGLHPEVARE 499
DB 776 ---HDANFQMLFMLMIGIVSLVXVQGITFAISSEKMYRANSQAFVILHODISFF 830
OY 500 -----GYFSALIA--PHLIGFELTVLLG--IVONPNIVNSV-----VALLSI 539
DB 831 DQOENTTGTALNTLSAGTKRELNG--ISGVTLITLIVSNVIVASIGVALVIGMKLALVCI 888
OY 540 AGV--LVGSGFLR 550
DB 889 SAVPALMGCFVR 901

RESULT 14
US-09-134-001C-3641
Sequence 3641, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779

```

```

PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3641
LENGTH: 273
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3641

Query Match
Best local Similarity 24.7%; Score 220.5; DB 4; Length 273;
Matches 62; Conservative 52; Mismatches 82; Indels 55; Gaps 7;

OY 66 RQILKDVSLVYESGQIMCILLSSGSGKTTLLDMSGRIGRAGTFLGEVYVNGRALRREP 125
DB 31 KHLLENINIKINGEFLAIVPGAGKSTILAKYI---LGLPQKEIIVDGKPFPGNKS 87
OY 126 QDCFSYVLQ-----SDTLSSLYRETLHYTALLAIRGNPSGFOKVEAVM 172
DB 88 SLKTSYVOKASAFNAGFPASVKEVYLSGLTKTKL-----FQRFNNNDVQ-KYTKVL 139
OY 173 AELTSHVADRLIGNSLGISTGERRRVSIAQLQDPKVMLEFDEPTTGDCMTANOIV 232
DB 140 ERLNLSHLT-----NNNIELSGGQOQRLIARALISDPSVVLDEPTNGIDAKHVSKEY 194
OY 233 VLLVELARRRRIYVLTIH-----QPRSELPQEDKIALISFGEIL 271
DB 195 DPLDKLAKKQGITIILVTHDIGVADTATEVACLKHLHFGSTEARKSIDEVEISK-----250
OY 272 IFQGTPAEMLD 282
DB 251 -IYGHPIQFVD 260

RESULT 15
US-08-665-259-27
Sequence 27, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSER: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/665,259
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1457 amino acids

```

TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-665-259-27

Query Match 6.68; Score 220; DB 3; Length 1457;

Best Local Similarity 32.18; Pred. No. 7.3e-14;

Matches 69; Conservative 37; Mismatches 91; Indels 18; Gaps 8;

```
QY 79 GOIMCIIGSSGSGKTTLLDMSGRIGAGTFLGEVYVNGRLARE--QPDCEFSYVLQSD 136
Db 1104 GBCFGLGVNAGKTSFFKMLTGDESTTG--GEAFVNGHSVLKDLLQVOQSLGCPQFD 1160
QY 137 TLSSLVVRETLH-YTALLAIRGNPGSFOKK-VEAVVAELSLSHVADRLIGNYSLGDIS 194
Db 1161 VPVDELFAREHLQYTRLRCT----PKKDEAQQVKKALEKLETKYADKPAGTY-----S 1211
QY 195 TGERRRYSIAQQLADPKVMEDEPTTGLDCMTANOIVLLAVELARRRRIYVLTTHOPRS 254
Db 1212 GGNRRKLSIAIALIGYPAFIFLDEPTTGMDPKARRFLMNLIDLITGRSVYVLTSHS-ME 1270
QY 255 ELFQLEFDKIALISFGELIFCGTPREMIDFENDCGY 289
Db 1271 ECEALCTRLAIMVNGRLHCLGSIQHLNRRGD-GY 1304
```

Search completed: June 11, 2003, 09:14:17
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:08:12 ; Search time 39 Seconds

(Without alignments)
3439.402 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MGDLSLTPGSGMKGLQVNRG.....PALVILGIYVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: SP_RHML_21:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	21.2	673	11 OBR543	08543 mus musculu
2	680.5	20.5	655	4 O96T8	096T8 homo sapien
3	672.5	20.2	655	4 O96L6	096L6 homo sapien
4	662	19.9	801	5 O8T691	08T691 dictyostell
5	660	19.8	657	11 O9R004	09R004 mus musculu
6	649.5	19.5	725	10 O9M3D6	09M3D6 arabidopsis
7	628	18.9	691	10 O8RW9	08RW9 arabidopsis
8	622	18.7	668	10 O9AR04	09AR04 oryza sativ
9	618.5	18.6	672	10 O9L182	09L182 arabidopsis
10	617	18.6	727	10 O9FNBS	09FNBS arabidopsis
11	615	18.5	692	5 P91892	091892 aedes aegypt
12	614.5	18.5	703	10 O8RXN0	08RXN0 arabidopsis
13	614	18.5	594	10 O9LJC3	09LJC3 arabidopsis
14	614	18.5	720	10 O9M2V7	09M2V7 arabidopsis
15	610.5	18.4	725	10 O9Z035	09Z035 arabidopsis
16	610.5	18.4	725	10 O9ASR9	09ASR9 arabidopsis

17	610	18.3	679	5 O9BH97	09BH97 ceratit c
18	608	18.3	708	10 O9M2V5	09M2V5 arabidopsis
19	602.5	18.1	654	10 O9L1W2	09L1W2 oryza sativ
20	600.5	18.1	670	5 O7V423	07V423 bactrocera
21	600	18.0	602	5 O9VC15	09VC15 drosophila
22	597	17.9	530	10 O9MAH4	09MAH4 arabidopsis
23	595.5	17.9	658	5 O16574	016574 caenorhabdi
24	595.5	17.9	687	5 O94960	094960 drosophila
25	595.5	17.9	785	4 O96L76	096L76 homo sapien
26	592	17.8	610	5 P90746	090746 caenorhabdi
27	591.5	17.8	740	10 O80946	080946 arabidopsis
28	589.5	17.7	646	10 O9C6R7	09C6R7 arabidopsis
29	588.5	17.7	648	10 O9C6W5	09C6W5 arabidopsis
30	587.5	17.7	627	11 O91WA9	091WA9 mus musculu
31	587.5	17.7	646	11 O9VBS9	09VBS9 mus musculu
32	584	17.6	755	10 O9ZUT0	09ZUT0 arabidopsis
33	583.5	17.5	687	10 O9C8K2	09C8K2 arabidopsis
34	581.5	17.5	687	5 O9NH94	09NH94 bombyx mori
35	580.5	17.5	669	5 O8WRP2	08WRP2 tribolium c
36	578	17.4	705	10 O9L1W1	09L1W1 oryza sativ
37	575.5	17.3	649	10 O9S1T6	09S1T6 arabidopsis
38	575.5	17.3	666	11 O9BEG9	09BEG9 rattus norv
39	573.5	17.2	669	5 O8WRP1	08WRP1 tribolium c
40	572	17.2	739	10 O9LFG8	09LFG8 arabidopsis
41	571.5	17.2	609	10 O9C8W6	09C8W6 arabidopsis
42	571.5	17.2	662	10 O949Y4	0949Y4 drosophila
43	571	17.2	609	5 O9VON4	09VON4 arabidopsis
44	567	17.0	626	5 O8T684	08T684 dictyostell
45	565	17.0	1528	5 O8T677	08T677 dictyostell

ALIGNMENTS

RESULT 1

OBR543 PRELIMINARY; PRT; 673 AA.

AC OBR543; 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sterolin 2.
GN ABCG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;
RT *Molecular cloning, genomic structure and characterization of novel
RT mouse head-to-head tandem ABC transporters.*;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF351811; AAL82898.1; JOINED.
DR EMBL: AF351799; AAL82898.1; JOINED.
DR EMBL: AF351800; AAL82898.1; JOINED.
DR EMBL: AF351801; AAL82898.1; JOINED.
DR EMBL: AF351802; AAL82898.1; JOINED.
DR EMBL: AF351803; AAL82898.1; JOINED.
DR EMBL: AF351804; AAL82898.1; JOINED.
DR EMBL: AF351805; AAL82898.1; JOINED.
DR EMBL: AF351807; AAL82898.1; JOINED.
DR EMBL: AF351808; AAL82898.1; JOINED.
DR EMBL: AF351809; AAL82898.1; JOINED.
DR EMBL: AF351810; AAL82898.1; JOINED.
SQ SEQUENCE 673 AA; 76008 MW; FA08340445DF259C CRC64;

Query Match 21.2%; Score 704; DB 11; Length 673;
Best Local Similarity 28.7%; Pred. No. 5.8e-43;
Matches 195; Conservative 130; Mismatches 261; Indels 94; Gaps 17;
11 GSKGLQVNRGSSSLGAPATPAPEPHSGILHASYSVSHRVR-----PAMD-I 57

```

Db      14 GTVADASGGLDSDL-----FSSESPNSLYFTYSGGSNTLEVDLTLYOVDIASQVWFEDL 69
Oy      58 TSCROOFTROI-----LKDVSLYESGQIMCIGSSGSKTLLDMSGRILRACGF 109
Db      70 AQFKIPWHSQSODCELGIRLSPKVSQGMALITGSSGCCRASLDVYTER-CHGGM 128
Oy      110 -LGEYVNGRALRREQFODCSYVLQSDTLSSLTVRETLHYATLLAI-RGNPGSPORK 167
Db      129 KSGQWINGOPSTPOLYRKCVAHVHQHOLLPLNLYRETLAFIAQMLPRFSQAOROKR 188
Oy      168 VEAVAAELSLSHVADRILGNLSIGISTGERRRVSIQAOLLODPVYMLFDEPTGLDCMT 227
Db      189 VEDVIAELRLQCCATRGVGMNTYKRGVSGGERRRVSIQVOLLMPICILLIDEPTSGLDSET 248
Oy      228 ANQIVLLVELARRRIVLVLTIHOPRSELPOLFDTIATISGELIFCGPAPMLDFPDC 287
Db      249 AHNLTITSLRKLAKGRNLVLISLHQRSDIFRLFDLVLTMTSGPTIYLGAQAOVMQYFSTI 308
Oy      288 GYPCPEHSPDFDYMDLTSVDTSKEREIETSKRVOMESAYKKA-----ICHKTLKN 341
Db      309 GPCPRYSPADFEYDLTSDIRSRKEREVAAYEKAQSLALELKVQGFDDFLMKAKE 368
Oy      342 IERMHLKTLPMVPRKTD-----PGVPSKGLVLRRTNLYRNKLAVITRLION 393
Db      369 LMTSTHYSLTL-----TQDTDCGTAVELPGMLEQSTLIRQISNDFRDLPTLLHGSBA 424
Oy      394 LIMGLIFLFLYLRVNSVLKGAIDORVGLIQFVQATPYTGMLNANVLEPYLRASVDES 453
Db      425 CLMSLIGFLYXGHAKOL--SFMDTALLFRIGALIFPNVILDVVSKCHSERSMUYEL 482
Oy      454 QDGLYOKMOMLAVLHVPSSVATMTSSVCYWTGLIHEVANRGYFSALLAPHLIG 513
Db      483 EDGYTAGPYFPFAKILGELPEICAVILYAMPYVLTMLRP-----VPELFL 529
Oy      514 EFLVLYLGIYONPIVNSVALS-----IAGVLYGS-----GFLRNIOEMPPIPKI 561
Db      530 LHLFLVMLVPCRCRMALASAMLETFHMSSEFCNALNSFYLAGFMINIDNMLIYAVM 589
Oy      562 ISYTFORKESEILVNEFYGLNFT--CGSSNVSTVTPMCAFTQGIQIEKTCGATSR 619
Db      590 ISKLSFLMKCSGLMOIQFNGHLYTTOIGNTFTSLGDTM-----ISA 632
Oy      620 FTMNFLIYSFTPALVIIGI 639
Db      633 MDLNSHPLYAYT--LIYIGI 650

RESULT 2
Oy      096TAB PRELIMINARY; PRT; 655 AA.
AC      096TAB;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
RT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
GN      ATP-binding cassette superfamily 6 (White) member 2.
OS      ABCG2.
OC      Homo sapiens (Human).
CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      NCI_Taxid-9606;
[1]
SEQUENCE FROM N.A.
RP      MEDLINE-21201983; PubMed-11306452;
RA      Komatsu H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,
RA      Arikawa H., Nishimura S.;
RT      "Identification of breast cancer resistant protein/mitoxantone
RT      resistance/placenta-specific, ATP-binding cassette transporter as a
RT      transporter of NB-506 and J-107088, topoisomerase I inhibitors with an
RL      indolocarbazole structure."
DR      Cancer Res. 61:2827-2832(2001).
DR      EMBL; AB051855; BAB46933.1;
DR      InterPro; IPR003439; ABC_transportr.

```

```

DR      InterPro; IPR003880; Pantene_attach.
DR      InterPro; IPR003580; Prochachylin.
DR      Pfam; PF00005; ABC_tran.1
DR      ProDom; PD000006; ABC_transportr; 1.
DR      SMART; SM00203; TK; 1.
DR      PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
DR      ATP-binding.
SQ      SEQUENCE 655 AA; 72314 MW; A8AF66B96034C5A8 CRC64;

```

```

Query Match      20.5%; Score 680.5; DB 4; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.9e-41;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

```

```

Oy      21 SSSLEGAPATAP---EPHSIGLIHASSVSHRVPMWDINSCROGMTROLIKDVSIVE 77
Db      13 SGNATNGPPATASNDLKFTBGAVALSFNNICRVALKSGFLPCRPRVERKILSNTGIMK 72
Oy      78 SGQIMCIGSSGSGKTTLLDMSGRILRACGFLEGVYNGRALRREQFODCSYVLQSDT 137
Db      73 PG-LNALIGPGGGKSSLLDYLARKPSG-LSGDVLLNG-APRANKKNSGYVODV 129
Oy      138 LLSLTVEETLHYTALLAIRGNPG-SFOKVEAVAAELSLSHVADRILGNYSIGISTG 196
Db      130 VMGTLVRENLOFSALRLATMTNHEKNERINRYIOELGLDKVADSKVTOFIRGVS 189
Oy      197 ERRRVSIQAOLLODPKVLFPDEPTGLDCMTANOIVLVLEARNRIVYLTIQPRESEL 256
Db      190 ERKRSTIMELITDPSILFLDEPTGLDSTANAVLLILKRRSKOGRTIITSHQPRSI 249
Oy      257 FQLEFKIALISFGEILFCGTPAEMIDPNDGYPCEPHSNPDFYMDLTSVDQO----SK 312
Db      250 FKLFDLSLLASGRMFHGPQOELGALYESGHCENANNDPFLDIINDSTRVALNR 309
Oy      313 ERE-----IETSKR---VOMIESANKSAICHT-----LNIERMKHLKTLPMVP 356
Db      310 EEDKATPEIIEPSSODKPLIEKLAIEYVNSFEYETKELHQLSGEKKKTIYFEKEISY 369
Oy      357 KTKSPGVSFKIVLIRVNTNLYRNKLAVITRLIONIMLFL--LEFVLRVNSVLK 414
Db      370 TT-----SCHOLRWVSKRSFNLGNPQASIAQITVYVLDLVGATFGKLNST----- 421
Oy      415 AIDRVGLIYQVQATPYTGMLNANVLEPYLRASVDESQDGLYOKMOMLAVL-HYLP 473
Db      422 GIORAGVLYF-FLTTNOCFSSVSAVELFVEYVERKLIHEIYISGYRVSSYFELGKLSLDLP 480
Oy      474 FSVVATMTSSVCYWTGLIHEVANRGYFSALLAPHLIGELTLVLGIYONPIVNSV 533
Db      481 RMPLPSIIFTCIVYFMGLKPKADAFVMMFTLM---MVAYSASSMALAIAAGOSVSV 537
Oy      534 VALLSIAGV--LVSGFLRNIOEMPPIPKIISYTFORKESEILVNEFYGLNFTCGSSN 591
Db      538 TLMTWICVPMIIFSGLLVNTLTIASMLSLQYFSIRYICRTALQHNHEFLQNFPCG--- 594
Oy      592 VSVTTPMCAFTQGIQIEKTCG 615
Db      595 LNTGNNPCNVA-----TCTG 610

RESULT 3
Oy      096LD6 PRELIMINARY; PRT; 655 AA.
AC      096LD6;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
RT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
GN      ABC transporter ABCG2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      NCI_Taxid-9606;
[1]
SEQUENCE FROM N.A.
RP

```


RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY017168; AAC52982.1;
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR003880; Pantine_attach.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transporter; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN.1.
 SO SEQUENCE 655 AA; 72288 MW; B3B5D0C02C095CA48 CMC64;

Query Match 20.28; Score 672.5; DB 4; Length 655;
 Best Local Similarity 29.08; Pred. No. 1.1e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SSSLEGAPATAP---EPHSLGILHASYSVSHRYRPMMDITSCROMTROILKDYSLVE 77
 DB 13 SSGTNGFPATASMDLKAFTGCAVLSFNITCYRKLSGELPCRKPEVEKILLNIMK 72
 QY 78 SGQIMCIISSSGSKTLLDMSGRAGTFLGEVYVNGRRLRQFODCFYVLOSDT 137
 DB 73 PG-LNAILGPTGGKSSLDVLAARKPSG-LSGDVLING-APPANFKCSGIYVDDV 129
 QY 138 LLSLYRETLHYTALLAIRGNPG-SFOKKVEAMELSLSHVADRILGNYSGISTG 196
 DB 130 VMGLTYRENLQFSALRLATYTMNHEKNERINRIQELGLKVAADSKVGQIFRGVSG 189
 QY 197 ERRRVIAIADLOPKMLFDEPTGDCMTANQIYVLVLAARNIVLTIHQPSSEL 256
 DB 190 EKKRTSGMELTDPSLLEPTGDSSTANAVILLLRMSKOGTITFISHQPRYSI 249
 QY 257 FOLEDKTALISFGLIFCGTAPAEMLDEFNDCGYPCPHSNPDEYMDLTSDVQ---SK 312
 DB 250 FKLEDSITLLASGLMHGPAQALGYFESAGYCEAYNNPAPDFLDIINGDSTAVANR 309
 QY 313 ERE-----IETSR---VOMIESAYKSAICHT-----LKNIRMKHLKTLPMVPF 356
 DB 310 EEDFKATEIIEPSKODPLLEKLAIEIVNSFYETAEHLQJSGCKKKITFYKISY 369
 QY 357 KTDSKPVSKLGLARVTRNLYRKNLAIVITLQNLIMGLFL--LEFVLRAVSNVLKG 414
 DB 370 TT-----SECHQLRWVSKRSRKNLGNPOASIAQIIVTVGLVIGAIYFGLKNDST--- 421
 QY 415 AIDRVGLLYQFAGATPYTGMLNANVLEFVLRAVSDQSDGLYOKNMMLAYAL-HVLP 473
 DB 422 GIORAGVLF-ELTNOCSESSVAVLEVEKLEFHEIYISGYRVSYFLGLSDLP 480
 QY 474 FSVVAAMIFSSVCYMTGLHPEVARFGYFSAALLAPHLIGEFLLVLIGIQNPNIVNSV 533
 DB 481 MRMLPSIIFTCIYFMHGLAKADAFVMMFTLM---MVAYSASSMALALAAQSVYSA 537
 QY 534 VALISTAGV--LYSGGLRNIOEMPIPKIISTFYFOKYCEILVNEFYGLNFTCGSSN 591
 DB 538 TLMTICFCYFMHFGILVNLITIASLWLOFISPIRYGTALQHNFIQGNFCPC--- 594
 QY 592 VSVTNPMAFTGIGIPIETKCPG 615
 DB 595 LNAITGNPCNYA-----TCTG 610

RESULT 4
 Q8T691 PRELIMINARY; PRT; 801 AA.
 AC Q8T691;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE ABC transporter Abcgl. 21, last annotation update)
 GN ABCgl.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 NCBI_TaxID=44689;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Aujard C., Loomis W.F.;
 RT "Evolution of the ABC Transporters of Dictyostelium";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF482380; AAL91485.1;
 SO SEQUENCE 801 AA; 90052 MW; CCC4F003CB195A3 CMC64;

Query Match 19.98; Score 662; DB 5; Length 801;
 Best Local Similarity 27.48; Pred. No. 8.3e-40;
 Matches 185; Conservative 134; Mismatches 246; Indels 110; Gaps 16;

QY 61 ROQWTRQILKNSLYVESGOIMCTIISGSGKTLTLLDMSGRAGTFLGEVYVNGAL 120
 DB 131 KKKISKQILTNINHIESGTFALMGPSGACKTLLDLARLHNLINS--GTMYLNGKS 188
 QY 121 RROQFODCFYVLOSDTLLSLVRETLHYTALLAIRGNPGSFQ-KKVEAMELSLSH 179
 DB 189 DENIFKICGYVQSDSLMPSLVRETLHYTALLAIRGNPGSFQ-KKVEAMELSLSH 248
 QY 180 VADLLIG--NLSLGISTGERRRYSIAAQLQDRKMLFDEPTGDCMTANQIYVLVE 237
 DB 249 CADTLVGTADNRKINGISGERRRYTISIELTGFSVILLDEPTGLDASTFYVMSALK 308
 QY 238 LARNRRLVLTIHOPRSELQDFPKIHLISFGLIFCGTAPAEMLDEFNDCGYPCPHSNP 297
 DB 309 LAKSGRTIITIHOPRSHITIDMPDNLILDGNTTYGKANKALEYFNANGYHCEKTNP 368
 QY 298 FDFYMD--TSVDPD-----TSKRVOMIESAYKKS--ALCHKTLKN 311
 DB 369 ADFELDLINQVEQADSDDDVDNDEEEIGGGGGSGCGAGTEDIGISIPMNSAV 428
 QY 312 ---KERIE-----TSKRVOMIESAYKKS--ALCHKTLKN 341
 DB 429 DNIRNNELKQDQ 488
 QY 342 IER-----MKHLKTLPMVPFETKNSPVFSLGVLARVTRNLYRKNLAIVITLQNLIM 396
 DB 489 ISKENRTDFYKTEK-----RGPNELTQFSLDLREYVNAHRHAPRVNLIQALIFQ 539
 QY 397 GLFELFVLRVRSVWLKGAIDRVGLYQFAGATPYTGMLNANVLEFVLRAVSDQSDG 456
 DB 540 G--LLCGIYVQQLGSSVSQSKGVAAFIIMGSPFAVNSTHVFDPVITIFLKDASG 597
 QY 457 LYOKNMMLAYALHVPFSSVAMIFSSVCYMTGL--LHPEVARFGYFSAALLAPHLIGE 514
 DB 598 VYDPLPEFLAKSPMDACIAVLPLPVATFIYMMNQNVRVDEPYSAAPFRFVLM---LVLA 654
 QY 515 FULVILGIQYON---PNI--NVSVALISTAGVLSGFLRNIOEMPIPKIISTFYFOKY 570
 DB 655 SQGCLSLGVLISSSVPWQVGTAVAPLIVILFELFSGFLINLNDVPMGLWFPYISFERY 714
 QY 571 CSEILVNEFYGLNFTCGSSNSVSTNPMCAFQIGIPIETKCGATSRFTMNLILYSE 630
 DB 715 MIEANVYNAKDVHFTLTDSD---KIGVCVPVOTGNNVIE--NMGYDIDHFRNVMILVLY 770
 QY 631 IPALVILGIYVFKIR 645
 DB 771 IIGFRVLTFLVILK 785

RESULT 5
 Q9R004 PRELIMINARY; PRT; 657 AA.
 AC Q9R004;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE Breast cancer resistance protein 1.
 GN ABCG2 OR BCRP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FVB: TISSUE-LIVER;
 RX MEDLINE-99413474; PubMed-10485464;
 RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.:
 RT "The mouse Bcrp1/Abcp gene: amplification and overexpression in
 cell lines selected for resistance to topotecan, mitoxantrone, or
 doxorubicin.";
 RL Cancer Res. 59:4237-4241(1999).
 DR EMBL: AF140218; AAD54216.1; -;
 DR MGI:1347061; Abcg2.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR003880; Pplantae_attach.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transport; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
 SQ SEQUENCE 657 AA; 73021 MW; 207870BC272CCD5 CRC64;

Query Match 19.88; Score 660; DB 11; Length 657;
 Best Local Similarity 28.08; Pred. No. 9e-40;
 Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

OY 13 MGIQVNRGSSQSLGAPAPAPPHSGILHASYSVSHRRPMDITSCROQTROLTKDY 72
 DB 12 MCRNRNNGLPNRNSRVRLAEADVLSFHIIYRV--KVSGLV--RTYKEKELISDI 66
 OY 73 SLVESGQIMCIIVSGSSGKTTLLDMSGRLGAGTFLGEVYNGALRREODPDSYV 132
 DB 67 NCIMKRG-LNALITGPTGGKSSLLDYLARKKPKG-LSGDVLNG-APPAHFGCCSGV 123
 OY 133 LOSDTLLSLVRETLHTALLAIRGNPGSPQ----KKEVAEMLSLSHVADRLNG 187
 DB 124 VDDVVMGTLVRENLQFSAALRL--PTTMKNEKNERINTIIELEKVAADSKVGT 179
 OY 188 YSLGISTGERRRVSIQAQLDOPKMLDEPTGTGCTAFANOIVLVLELRNRNIVLV 247
 DB 180 QTRIGISGGERKRTSIGMELITDPSILFDEPTGTGDSSTANAIVLLLRMSQGTIIF 239
 OY 248 TIHPRSLEFQDLFKIAILSPGELIFCGTAPAEMLDFNDCGYPCEHSPNPFYMDLTGV 307
 DB 240 SIHPRYSTIFKLFDSLTLASGLYVHGPAKALEFASAGHCEPYNNPADPFLDIVNG 299
 OY 308 DPGS-----KEREIEFSKR-----VOMIESAYKKAICHTKLNIERKKILKTL 352
 DB 300 DSSAVMLNREEDNENANKTEEPSKGEKPYENLSEFYINSAIYGETKALDOL----- 352
 OY 353 HVPFKTDSGVFSKLV-----LARRVRNLRNKLAVITRLONL 394
 DB 353 -----PGAOEKKGTSAFKEPYVYVSFCHQLMRNARRSRKLNLGNFQASVAOLITV 403
 OY 395 IAGLFL--LEFVLAVRSNVLKGAIDRVGLLYQVGAAPPTGMLNANVLEPVLRAVSDE 452
 DB 404 ILGLIIGAIYFDLTKDA---AGMNRAGVLF-PLTNGCCFSVSANVELFYVEKILFHE 458
 OY 453 SDDGLYOKROMLAVAL-HVLPFSVAATMIESVCTWTGLHPEVARFGYSALLAPHL 511
 DB 459 YISGYRVSSTYFEGVMSDLEPMRFLPSYIFCIILYFMGLTKTDAEFTIIMFTIL--M 515
 OY 512 IGEFTLVLTGIVONPNIYNSVALLSTAGV--LVGSGFLRNIOEMPIFKIISTPTOK 569
 DB 516 VAYTSSAMALAIATQSVAATLTMTLTAFFMFGSLVLRIGTGLWLOQFESIPR 575
 OY 570 YCSEILVYNEFYGLNFTCGSSNVSTINPMCAFTGOGIOFIERTKCG 615
 DB 576 YGFTALQYNEFLGQFCPG---FNTDNTSCVNSAI-----CTG 612

RESULT 6
 ID 09M3D6 PRELIMINARY; PRT; 725 AA.
 AC 09M3D6;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ABC transporter-like protein (Putative ABC transporter protein).
 GN T26112.10 OR A3G55130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid-3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Monfort A., Casacuberta E., Puigdomenech P., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Queller F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carlinici P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Kosemura E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.M.,
 RA Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T26112.10/A3G55130 (GI:7019646).";
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Eguu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL132954; CAB75747.1; -;
 DR EMBL: AY045932; AK76606.1; -;
 DR EMBL: AY079387; AA185118.1; -;
 DR InterPro: IPR003593; AAA_Atpase.
 DR InterPro: IPR003439; ABC_transport.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR ATP-binding.
 KW KW
 SQ SEQUENCE 725 AA; 80656 MW; 790C535A7929CC16 CRC64;

Query Match 19.58; Score 649.5; DB 10; Length 725;
 Best Local Similarity 29.48; Pred. No. 6e-39;
 Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

OY 33 PEPHSILGILHASYSVSHRRPMDITSCROQTROLKDVSLVESGQIMCIIVSGSSGK 92
 DB 68 PVYVYLNFNNDLYDTLRR-----FGFSQNRNVTKLDDVSEADCDITAVLGASGAK 123
 OY 93 TILLDMSGRLGAGTFLGEVYNG-RALRREQFODCSYVLAOSDTLLSLVRETLHT 151
 DB 124 STLDALAGRYAE-GSLGSYTLNGEKVLAQSLKLVISAYVQDDLLPMLTKETLWFA 182
 OY 152 ALLAIRRG-NGSPQKYEAVMAELSLSHVADRLGNLSLGISTGERRRVSIQAQLD 210
 DB 183 SEFRLPRLSLSKMKEREALDLDGLNRNANTVYIGDHGHSVSGGERRRVSIIGDIIHD 242
 OY 211 PKVMLDEPTGTGCTAFANOIVLVLELARRNRIVLTIPRSELPOLFKRIALISFGE 270
 DB 243 PIVLEDEPTGTGDSSTANFNVVYVYLAIRNAGSGSYIMSIHRPSARIYELDRILITLSNGK 302
 OY 271 LIFGCTPAEMLDFNDCGYPCEHSPNPFYMDLTGVSVDGSKEREIETS----- 319

```

Db 303 SVENGSPASLPGEFSDGRIPEKENISERAILDV-----RELEGSNEGTAALVDN 354
QY 320 -----KRWOMIESAYK-----KSAICHKVL--KNIERMKHLKTLPMVPEFKTD 360
Db 355 ERMQONKISLIQSPQNKLDODRSLSLKEAINASVGRGLVSSSSNPTSEVSYA 414
QY 361 SPVEFSKLVLRVTRNLVKNLAVITRLONLIMGLFLLFVLRVSNVLCGAIDRV 420
Db 415 NPSLFEFE-ILARRKMKMIRMBELVOTRIATVAVTGC-LLATVYMKLDHPTPGA-QERL 471
QY 421 GLLYOFGATPYTGMLNANVLFPVLRVSPQESODGIXOKWOMLALALVLPESVATM 480
Db 472 -TLFATVYPTMFCCLDNVPEFIOERTIFURTHNAVRYSSVYISHLSVLPQLAPSL 530
QY 481 IFSSVCWTGLGLHPEVAREGFSAALAPHLIGEFLTLVLGIYONNIYNS--VVALLSI 539
Db 531 VFSATIFMTVGLSGGLEGFVYCLLIYASFSSSVTFISGVY--PNMLCYMVSITTYL 588
QY 540 AGVLVSSGFLRNIOEMPIPKIISYTFPOKCEILVNEPFGMLFTCGSSNVSVTTPM 599
Db 589 ACCLISGFVYVNDRIPEYTWMPHYISILKPYEAVLINEF-----DDPS 633
QY 600 CAPTGOIOFEKTPGATS 618
Db 634 RCFVKGVOYVFDSTLLGVS 652

RESULT 7
Q8RM19 PRELIMINARY: PRT: 691 AA.
AC Q8RM19; 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Hypothetical 77.2 kDa protein.
GN AT3G21090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carinl P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY093054; AAM13053.1;
DR Hypothetical protein.
KW
SQ SEQUENCE 691 AA: 77219 MW: CEA73CC0B440D7E9 CRC64;

Query Match 18.9%; Score 628; DB 10; Length 691;
Best Local Similarity 28.1%; Pred. No. 2,1e-37;
Matches 173; Conservative 123; Mismatches 225; Indels 94; Gaps 17;

QY 25 LEGAARLAPE-PHSGILHASYSVSHRVPMDITSCROOQ---TRQILKQVSLYVESG 79
Db 3 LEGSSSGRRQLPSKLEMSRGAYLA-----WEDLVVYVJPNFSDGPTRIQLQNGYAEPG 56
QY 80 QIMCIGSSGSGKTTLLDMSGRGRAGTFLGEEVYVNGRALREDFQDFSVILOSDTLL 139
Db 57 RIMALMGPSGSKSTLLDSIAGRRLARVYMTGNLNGKALD--YGLVAVTQEDVLL 114
QY 140 SLVTRFETLHYTALLAIRGNGSPQOK-----VEAVNAELSLSHVADLIGNSLGIGIS 194
Db 115 GLTVRETIYSAHLRL-----PSDKSKEVSDIVSGTIEELGLDQCSDRVIGNMARGIS 170
QY 195 TGERRRVSIAMLODDPKVLMFDEPTGIDCFTANQIYVLLVELARNRIVVLTTHQPS 254
Db 171 GERRRVSAILEILTRPQILFLDEPTSGIDSASAFVIOALNINARDGTVISVHOPSS 230

```

```

QY 255 ELFOLFDKIALISFGLIFCGTDAEMDFENDGCPQCEHNSPDPFMDLTSVDOSKER 314
Db 231 EVRLFDLDLGLSSGESVTFGEAKSAVEFPAESGFPCKPRKNSDHLRCINDPDTVA 290
QY 315 EIETSKRVQ-----NIESAYKSAICHKTLKNIEMKHLKTLPMV 354
Db 291 TLKGSRIQETPATSDPIMNLATSVIKARLVEN-YRKSRAKSAKSNIRELSNIEGLE 349
QY 355 PEKTSQGVFSKLVLRVTRNLVKNLAVITRLONLIMGLFLLFVLRVSNVLCG 414
Db 350 IRKGSERATW-WKOLRLTLARSFYIMCHDVGYWTRISYIVSI----- 392
QY 415 AIDRGLGLXOFGATPYTGMLNANVL-----FVL--RAVSDQESODG 456
Db 393 ----SVGTIFYDGYIS-YTSLIARVSCGGFTGEMFMSIGGFPSPLEEMKVFYKERTSG 447
QY 457 LYOKWOMLALALVLPESVATMTIFSSVCWTGLGLHPEVAREGFSAALAPHLIGEFL 516
Db 448 YGVSVYILSNYISSPFLVAISVITGTYINLYKRRPFSHYAFLCLNIFSVSYESL 507
QY 517 TLVLGIYONPNIVNSVALLSIG-VLVSGFLRNIOEMPIPKI-----ISYTFQKY 570
Db 508 MMYVASVY--PNFLMGLITGAGLIGIIMNTSGFRLPLDP--KIFMRYPVYSIGSW 562
QY 571 CSRLVYVNEFYGNF 585
Db 563 AIQGYKNDPLGLEF 577

RESULT 8
Q9ARU4 PRELIMINARY: PRT: 668 AA.
AC Q9ARU4; 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Putative ABC transporter.
GN P0445D12.3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. NIPONBARE;
RA Sasai T., Matsumoto T., Yamamoto K.;
RA "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0445D12.3";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AP003046; BAB40032.1;
DR InterPro: IPR003593; AAA_Arase.
DR InterPro: IPR003439; AAA_Transport.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Transport.
KW
SQ SEQUENCE 668 AA: 73368 MW: D1875B8C75B0F3B2 CRC64;

Query Match 18.7%; Score 622; DB 10; Length 668;
Best Local Similarity 31.0%; Pred. No. 5,4e-37;
Matches 171; Conservative 100; Mismatches 245; Indels 36; Gaps 12;

QY 55 WITSC-----RQMTROLKQVSLYVESGQIMCIGSSGSKTTLLDMSGRGRAGT- 108
Db 58 WARTICALKNKRGDVARELLSNASGEAKSRLALMGPSGSKTLLANVLAQOLTRASPSL 117
QY 109 -FLGEEVYVNGRALREDFQDFSVILOSDTLLSLTVRETIYTAALLIRRG-NGSPQK 166
Db 118 HLSGFLYINGRIRISGGRK--IAYVROEDLFFSOLTVRETIYSLAELOLRRLTPERRS 175

```

	RESULT 9		
Q0	09L182		
ID	09L182	PRELIMINARY;	PRT; 672 AA.
AC	09L182;		
DT	01-OCT-2000 (TEMBREL. 15, Created)		
DT	01-OCT-2000 (TEMBREL. 15, Last sequence update)		
DT	01-JUN-2002 (TEMBREL. 21, Last annotation update)		
DE	ABC transporter-like protein.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OX	NCB1_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-COLUMBIA;		
RA	Kaneke T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-COLUMBIA;		
RA	MEDLINE-203630399; PubMed-109707553;		
RT	Nakamura Y.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. II:		
RT	Sequence features of the regions of 4,251,695 bp covered by ninety P1,		
RT	TAC and BAC clones."		
RL	DNA Res. 7:217-221(2000).		
EMBL	AB001313; BAB03081.1; -		
DR	InterPro: IPR003593; AAA_Arpase.		
DR	InterPro: IPR003439; ABC_transport.		
DR	InterPro: IPR003880; Pfam_attach.		
DR	Pfam: PF00005; ABC_tran; 1.		
DR	ProDom: PD000006; ABC_transporter; 1..		
DR	SMART: SM00382; AAA; 1.		
DR	PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.		
DR	PROSITE: PS00012; PHOSPHOTANETHEINE; UNKNOWN_1.		
ATP-binding.	672 AA; 75269 MW; 20B2D99215600135 CRC64;		
Q0	SEQUENCE		

```

Query Match      18.6%; Score 618-5; DB 10; Length 672;
Best Local Similarity 32.1%; Pred. No. 9.8e-37;
Matches 182; Conservative 10%; Mismatches 206; Indels 79; Gaps 17;

OY    66  RQILKDVSLVYESGQIMCILLSSSGSKRTLLDMSGRGLRGAGTFLGEYVNGR-----ALR 121
       | : || | | : | :: : || | || || || : || || | : || :
DB    97  RLVLKCVSIGVYPGELLMLCPGGSGCKTTLVTALAGR- -OGKLSCGVSYNGEPTSSVK 154
       | : || | | | : | : | : | : | : | : | : | : | : | :
OY    122 REOFODCRSVYLQSTLLSILTFRLETHYTLLAIRGNPSPQK-----VEDVAMLS 176
       | : || | | | : | : | : | : | : | : | : | : | : | :
DB    155 RKT-----GFYTQDDVLYPLHLYMETLTYTLLRLK- -PELTRKEKLEDVEAVSDLG 205
       | : || | | | : | : | : | : | : | : | : | : | : | :
OY    177 LSHVADRLLIGNSLTGISTGERRRVSIAQLLODPKVALLFDEPTGLDCMTANOIVLLV 236
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    206 LTRCNVSIVGGGLIRGISGERKRYSIGEMLVNPSLIILDEPTSGLDSTAAIRIVATLR 265
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    237 ELARNKRIYVLTINHPRELFQLEPKMILISFGELIFCGITPAEMLDFFNDCCG-PCPEHS 295
       | | | | | | | | | | | : | : | : | : | : | : | : | :
DB    266 SLARGGRVTTTIHQPPSSRLYMFPKVLVLSGEGPIYSGDGRWMEYFGSIGYOPGSGSFV 325
       | | | | | | | | | | | : | : | : | : | : | : | : | :
OY    296 NPFDYMDLTVDNQSKER--EIEFSRKVMQI-----SAKKSAICHTIKNIIE 343
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    326 NPADVELDLANGITSIDTKYOTIQETINGHLDRLEONSNYSKOSLISYKN--LYPLEA-- 381
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    344 RKMKHLKTLPMVPFKTKDSPGV-----FSKLVLLRYVRNLVNRKLNAVIT 388
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    382 --EVSRTEFP-----QQDTNMRLRKRAKATTNNWPISMWGFSVLAK---RGLAKESHESFS 430
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    389 RLNLNLINGLFLFEVLVAVRSNVLKGAIQDRVGLLYOFVGAPPTYGMILNAVLPVILAV 448
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    431 GLRIFMVASVSLISGLIMMHRSRV-AHLQDOVOGLLFFFSIPMGFPFLFNATFEPQERPM 488
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    449 SDOESQDOLYOKOMQMLLAVALHYLFESVVAAMISSVCYWTGLGHPEVARPGYSALLA 508
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    489 LIKESSTSIYNKRSITYIARTYGDDIPMELIIPFTVITTYMMGKPSLTFITMITLMIVLY 548
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    509 PHLIGEFTLVLLGIQNPNIVNSVALLSIAGLVGSGFRNIQENPIPKITSYTFQ 568
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    549 NVIVAQGAGLAIAMDARKAARLLSVSLMVFLLAGY----IQHIPGLIAMLKIVSFS 605
       | : | : | : | : | : | : | : | : | : | : | : | : | :
OY    569 KYCEIILVNNEFYGLNT-----CGS 589
       | : | : | : | : | : | : | : | : | : | : | : | : | :
DB    606 HCYKILV-----GYQYTWDEVYECGS 627
       | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10
OGENB5 PRELIMINARY; PRT; 727 AA.
AC OGENB5;
DB 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ABC transporter-like protein.
DS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones." ;
RL DNA Res. 4:291-300(1997).
DR EMBL: AB006704; BAB08684.1; -
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF000005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
```

DR SMART: SM00382; AAA. 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 KW ATP-binding.
 SQ SEQUENCE 727 AA; 80300 MM; 63450CF587213A4 CRC64;

Query Match 18.6%; Score 617; DB 10; Length 727;
 Best Local Similarity 28.7%; Pred. No. 1.4e-36;
 Matches 179; Conservative 126; Mismatches 233; Indels 86; Gaps 19;

QY 35 PHSIGILHASYSVSHRVRPMW-----DISCROQWTRQLIKDVLVESGOI 81
 D 65 PVLSTFDLRTYSVVRKFTWRBSVSSDPAPSGITSSK---TKILNGITGARDEI 121
 QY 82 MCILGSSGKTTLLDMSGRLGAGTFLGEVYNGRALRRPODFSVYLSQDTLLS 141
 D 122 LAVLAGSGSGKSTLIDALANRIAK-GSLKGNVTLNGEVLNSKMKKSAIVYMDDLLFPM 180
 QY 142 LTVRETLHYTALLAIRGNPGSFOK-----KVEVMAELSLSHVADRLLIGNYSIGISTG 196
 D 181 LTVRETLHYTALLAIRGNPGSFOK-----KVEVMAELSLSHVADRLLIGNYSIGISTG 196
 QY 197 ERRRSVIAIADLDPKVMLEDEPTGLDCTANOIVLVLELARNRIVLTIHQPSSEL 256
 D 237 ERRRSVIGIDILHDPILFLDEPTSGIDSTALSIVIKRLKIOGSMVITLHQPSYRL 296
 QY 257 FOLEDKIALISFGLFCGTPAEMLDFFNDGYPCEPHSNPFDFYMDL-----TSYDTQ 310
 D 297 LRLDLRLFLRSGQTVSGSPALMPREFAFGHPHPEHNRTEFALDLINELSGAGSTR 356
 QY 311 S-----KEKELETSKRVQM-IESAYKKSALCHTKLNIERMKH-----LKTLPV- 354
 D 357 SLVEFNKGFORAKAPRSQGLSLKEAISASISKGLVSGATTTTHSSGSPVSTIPTFA 416
 QY 355 -PEKTKDSPVFSGLVLRVVRNLRNKLAVITRLQNLINGLFL--LEFVLRAVSNV 411
 D 417 NPF-----WEELAVLAKRSMTNSRQPELFGIRGAVLVGTFILATMEWOL---DNS 465
 QY 412 LKGAIDRVGLVQVFGATPYTGMLNVLNLEPVLRVAVSDQSDGLYQKQOMMLAYALHV 471
 D 466 PKG-VQERIG-CFAFAMSTFFTCADALPVFLQERFLPMRETAVMNARRSYVLSHVA 523
 QY 472 LPSVAVTMIFSSVCYWTGLLHPEVARFGFSAALLAPHLIGELVLVLGIYQNPV- 530
 D 524 LPSLITLISLAFAITFGVGLDGLMGFLFYFLVILLASFAGSSFTVFLSGV--PVM 581
 QY 531 --NSVVALSLIAGLVGSGFLRNIOEMPIPKIISYFTPOKYSCELLIVNEFGLNCTG 588
 D 582 GYITVAAIL--ATFLFSGFFINRDRIPGIWIMFHTISLVKPYEAVLNEF----- 631
 QY 589 SSNVSVTTPMCAFTGQIOFIETKT 612
 D 632 -----GDPTKCFVRGQVQIFDNT 648

RESULT 11

P91892

PRELIMINARY; PRT; 692 AA.

AC P91892;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE EYE pigment TRANSPORTER (EYE pigment transporter).
 GN WHITE.
 OS Aedes aegypti (Yellowfever mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Aedes.
 NC NCBI_TaxID=7159;
 RN (1)
 RM SEQUENCE FROM N.A.
 RC STRAIN=ROCKEFELLER;
 RX MEDLINE=97418473; PubMed=9272447;
 RA Coates C.J., Schaub T.L., Besansky N.J., Collins F.H., James A.A.;

RT "The white gene from the yellow fever mosquito, Aedes aegypti."
 RL Insect Mol. Biol. 6:291-299(1997).
 RN 121

RP SEQUENCE OF 1-20 FROM N.A.
 RC STRAIN=ROCKEFELLER;
 RA Coates C., Schaub T.L., James A.A.;

RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; U88851; AAC04894.1; -
 DR EMBL; AF051096; AAC05165.1; -
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transport; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 SQ SEQUENCE 692 AA; 77231 MM; DF2F0EBB4C17641 CRC64;

Query Match 18.5%; Score 615; DB 5; Length 692;
 Best Local Similarity 27.3%; Pred. No. 1.8e-36;
 Matches 169; Conservative 135; Mismatches 226; Indels 88; Gaps 17;

QY 66 RQLIKDVLVESQIQCILGSSGKTTLLDMSGRLGAGTFLGEVY--NGRALRR 122
 D 110 KHLKNTGMAKSGSELLAVNGSGAGKTTLLNLSFR-SPGVKIAPTSVALNGIPVNA 168
 QY 123 EOPDCSVYLSQDTLLSLTVRETLHYTALLAIRGNPGSFO-KVEVMAELSLSHVA 181
 D 169 EQLARCAIVQDDLFIPALTREHLVFMALNRGKVPKSVKNNRNEVQLSELAKCA 228
 QY 182 DRLIGNS-LGSGTGERRSVIAIADLDPKVMLEDEPTGLDCTANOIVLVLELAR 240
 D 229 DTIIAGARKMGKLSGGERKTLAFASELITPPLLIDEPISGIDSFPAHSLVQLKGMAL 288
 QY 241 RNRIVLTIHQPSSELFOLEDKIALISFGLFCGTPAEMLDFFNDGYPCEPHSNPFD 300
 D 289 KGRITITLTIHQPSSELCLEDKILLVAGHVAFLGSPYQASEFSGIICPPRYNPAD 348
 QY 301 YMDLTSVQSKKELETSKRVQMIESAYKKSALCHTKL-----KNIEMKHLKTLPMV 354
 D 349 YVOMLAI--APKNKACRPTIKTKIDCFVAVSPARVPMVANSGRNVEQYTLQ----- 400
 QY 355 PFKTKDSPGV-----FSKLGVLRLRVNLRNKLAVITRLQNLINGLFL--LEFVLRAV 408
 D 401 PMEGASMTGYRSMWQFYVLRMSMTVLAKDMLVKNRLLQRMATLIGSITFGRLD 460
 QY 409 SNVLKGAIDRVGLVQVFGATPYTGMLNVLNLEPVLRVAVSDQSDGLYQKQOMMLAYA 468
 D 461 QD-----GVNNINAGLFLFTNMTFQWVFAVINVESALPVFLREKRSRLRVDYFLGKT 516
 QY 469 LHVLPVSVA TMIFSSVCYWTGLLHPEVARFGFSAALLAPHLIGELVLVLGIYQNP 528
 D 517 IAEVPLELAVPEVFTSTITPMIGLKSAT--YTLALTLIVVA-----N 559
 QY 529 IVNSVVALSLIAG-----VLVSGFLRNIOEMPIPKIISYFTPOKYS 572
 D 560 VATSPGLISLCASSISMALSVGPVLIIPRLIGGFLNLSASVSEFVISTYSWRPYAN 619
 QY 573 ELIVNEEFTGL--NFTCGSSNVSVTTPMCAFTGQIOFIETKCP-GATSRFTMFLIL 627
 D 620 EALLINQMSVQDEGIACITRANV-----TSPSSGQIILETFNFKVE 660
 QY 628 -VSFIALVILGIVPEKI 644
 D 661 DFGFDIACDMLTIVIFRL 678

RESULT 12

O88XNO

PRELIMINARY; PRT; 703 AA.

AC O88XNO;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

Query Match	18.5%	Score 614.5;	DB 10;	Length 703;
Best Local Similarity	29.4%	Pred. No. 2e-36;		
Matches 170;	Conservative 112;	Mismatches 233;	Indels 63;	Gaps 16

RESULT 13	AD	PRT;	594 AA.
09LJC3	09LJC3		
AC	PRELIMINARY;		
09LJC3			
DT 01-OCT-2000	(PRELIMrel. 15,	Created)	
DT 01-OCT-2000	(PRELIMrel. 15,	Last sequence update)	
DT 01-JUN-2002	(PRELIMrel. 11,	Last annotation update)	

Query Match	18.5%	Score 614	DB 10	Length 554
Best Local Similarity	28.0%	Pred. No. 1.8e-36		
Matches 173	Conservative 123	Mismatches 229	Indels 92	Gaps 17

```

QY 25 LEGAPARAPE-PSHLGILHASTYSVSHRVRPMMDITSCROOM---TROILADVLSLYESG 79.
Db 3 LEGSSSGRRLOPSPKSLKMSRGATYLA-----MEDITLYIPNFSDDGTTRILRLRLNMGAYEPG 56
QY 80 QIMCITLSSSGSGKTTLLDAMSGRLGRAGTGFGEYVYNRRLRREDFOCFSYLOSPDLL 139
Db 57 RIMAMGSSSGSKTLLDSDSLAGRLARNVMTGNLLNKKARLD--YELVATYVQEDVLL 114
QY 140 SSLTRETLMATTALLAIRKGNPFSFOKK-----VEAVMSELSHVADRLGNISLGIS 194
Db 115 GTLVRETITYSAHLRL---PSMSKEEVSIDIYEGITMELGLODCSDRVIGNHARGVS 170
QY 195 TGERRRVSIANAOLLODPKVMLEDPPTGLDQMTANOIVYLLVETLARRRRIYVLTIHORS 254
Db 171 GGEKRRVSIAMELITRPOILLDEPTSLGDSASAFYIOALRNARROGRITYISSHOSS 230
QY 255 ELFOLEFKIALISFGELIFCGTPAEMLEDFENDCQPCPEHSPNPFYMDLTSVDOYSKER 314
Db 231 EYFALFLODFLTLSSGESYVFEAKSAVEFSESGPCPKRKNPDSOHLFRLCINSDFDTYA 290
QY 315 EIENSKRVO-----MIESYKKSALCHTKLKNIERKKHLTKLPWY 354
Db 291 TLKOSQRIQETPATSDPLMANLANSYIKARLVEN--YKRSKYAKSAKSRIRELSNTEGLEME 349
QY 355 PFKKSDPGVFSKLGVLRLRYTRNLYVNKLAVITRLLONLIMGFLFLFYLVRSNVLKG 414
Db 350 IRKSESEATW-WKQRLTLTARSLFMCMGDGVYMRITSIYVSL----- 392
QY 415 AIODRVALLQFVGATPYTGMLANVL-----FPVL--RAVSDQESODG 456
Db 393 ----SVGTIYDYGVS-YTSLILAVSOGGFTYGMTPMSJGGFSFLEMKVFEKERLSG 447
QY 457 LYOKOMMLLAVLHVLPFSVAATIFSSVCWMTGLHPEVARGEFYSALLAPHLIGEL 516
Db 448 YGGSVYILSNYSFPELVATISYTGITITYNLVKFRPGSHVAFCLNITPFSVSYESL 507
QY 517 TLVVLGIVQNPJNVNSVVALISLIG-VLVSGGFLRNIQEMP----IPKIIISYFTQOKY 570

```

Db 508 MAAVAVSV--PNIAGLITGAGLIGITMTSGFFRLDLPKIFWRRPVSYISYQMAIQ 565

QY 571 CSEILVNEEYGLNFTC 587

Db 566 AS-LRSGSKAGESRTC 581

RESULT 14

Q9M2V7 PRELIMINARY; PRT; 720 AA.

AC Q9M2V7;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE ABC transporter-like protein.

GN T1SC9.80.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI_TaxID=3702;

OX [1]

RP SEQUENCE FROM N.A.

RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,

RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quettier F.,

RA Salanoubat M.,

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL132970; CAB82704.1;

DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.

DR Pfam: PF00005; ABC_tran; 1.

DR ProDom: PD000006; ABC_transporter; 1.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOMN_1.

DR ATP-binding.

SO SEQUENCE 720 AA; 79889 MW; 9D16431098CF1A3 CRC64;

Query Match 18.5%; Score 614; DB 10; Length 720;

Best Local Similarity 28.7%; Pred. No. 2,3e-36;

Matches 182; Conservative 122; Mismatches 254; Indels 76; Gaps 17;

QY 23 SSLEG--APATAPEPHSIGLTHASYSVS-----HRRVRWMDITSCROOQTQOILKDV 73

Db 40 SSIDGNDHLMRPVPPVLSFNNLTYNVNRKIDFDHVLVWRRFSFK--TTLLDNIS 96

QY 74 LYESGOIMCIISSSGSKTTLIDAMSGRLGRAGTEGEEYVNGRLRREQFODCSYVL 133

Db 97 GETRDEILAVLAASSGSKSTLIDALANRIAK-GSLKGYTLANGELQSMKLVISAYVA 155

QY 134 QSTLLSSGLTVRETLHTTALLAIRKNPGSGFQ-KYEAVALSLSHVADRLIGNYSLG 192

Db 156 OODLFPMLTVEETLFAAEFRILPRSLPKSKKRLVQALIDQIGIRNAATIIIGDSHG 215

QY 193 ISNGERRRYSIAQQLODPRVLMFDEPTGLDQMTANQIVLVYLAKRRRYVLTIHOP 252

Db 216 ISGERRRYSIGIDIIHDPVLFLEDEPTGLDSTSAFMYVVKLRIASGSIITIMSIHOP 275

QY 253 RSELPQLFKIALISFGEILFCSTPAEMLDFFNDGCPCEPHSNPDEYVADLTS----- 306

Db 276 SHRYVLSLDRILIFLSGHYFSSPASLPSFAGFNPPIPENNOEPALDILRELGSA 335

QY 307 -----VDTOSKREIETSKRVQIMSAKKAICKRTIK-----NIERKKLK----- 349

Db 336 GGTGRIYEFNKKQWKKOSNPQTLTPPASPNP--NLTLKEAISISGKLVSGGGGS 393

QY 350 -----TLMPWPFETKDSFGVSKGLVILRRVYTNLVNKNKLAIVTTLQNLINGLFEL 401

Db 394 SVINHGGLTAVAFANP-----FWIEIKTLTRRSILNSRQPELLQMLATVIVTG-FLL 448

QY 402 FFVLKRVSNVLKGAIDORVGLIXOFVATPYTGMILNVLFPVLRASVDSQSDGLYOKW 461

Db 449 ATFWWRIDNSPKG-VQERLG-FEFAFNSTMYTCADALPVLQRIYFMRETAANARRS 506

QY 462 QMLAVLHVLPESVAVTMIFSSVCYVTLGLHPEVARNGFSALLAPHLIGELTVLV 521

Db 507 SYVLSHAIIVTFPSLIFLSLFAVATTFWAVNGLEGIMGFLFCLIIILASFMSGSSEFTLS 566

QY 522 GIVONPNIV--NSVALLSLAGVSGFLRNQOEPIPKIISYTFPOKCEILVYN 578

Db 567 GVV--PVMGLGITIVAIL--AFLLSGFEFINDRIPQYIMFHYLSLVKPYEAVLON 622

QY 579 EFGYLNFTCGSSNVSVTNPMCAFTQIGIETK 612

Db 623 EF-----SDPEECVRGVQLFDNS 641

RESULT 15

Q9ZU35 PRELIMINARY; PRT; 725 AA.

AC Q9ZU35;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative ABC transporter.

GN ATG01320.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI_TaxID=3702;

OX [1]

RP SEQUENCE FROM N.A.

RA STRAIN-CV. COLUMBIA;

RX MEDLINE-20083487; Pubmed-10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Salzbarg S.L., Fraser C.M., Venter J.C.;

RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana.";

RL Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA STRAIN-CV. COLUMBIA;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

CC 1-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL: AC006200; AAD14532.1;

DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.

DR Pfam: PF00005; ABC_tran; 1.

DR ProDom: PD000006; ABC_transporter; 1.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.

DR ATP-binding; Transport.

SO SEQUENCE 725 AA; 78899 MW; 7DB2E556FE3553D7 CRC64;

Query Match 18.4%; Score 610.5; DB 10; Length 725;

Best Local Similarity 29.7%; Pred. No. 4,1e-36;

Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

QY 51 VRP-----WMDITSC-----ROOQTRQILKDVSLVSGSGLTLLDAMSGR 102

Db 65 IRVYTRKNRNTITSLSDSKSKYRPLKKNVSGAKGRRILALNGPSSGKTTLLNVLAV 124

QY 103 LGRAGT--FLGEYVNGRRLRREQFODCSYVLSQSDTLSSLTVRETLHTTALLAIRR-G 159

Db 125 LSLSPRLHLSGLLEVNGKRPSSKAYK--LAFVQEDULFSQLVRETLSPFAELQLEIS 182

```
OY 160 NPGSFQKVEAVVAEISLSHVADRLIGNYSLGSGISGERRRVSIANAQLQDPKVMLEDER 219
Db 183 SAERREYVNNLLKLGVLSCADSCVGDNAKVRGISGGERKRLSLACELASPSVIFADEP 242
OY 220 TTGLDCMTANOIYVLLVEELARRRRIYVLTIHQPRSELFQLEFKIATILSPGELIFCGTPA- 278
Db 243 TTGLDAFQAEMETQKLAQDGHYICSIHQPRGSYAKFEDIYLLTEGTLVYAG-PAG 301
OY 279 -EKLDFFNDCCGIPCEPHSNPFDPYMDLTSTVDPTQSKEREIETSKRVOMIESAY--KKSATC 335
Db 302 KEPLTFEGNFGELCPEHNPDAEFLADLISVDYSSSEYVSSQKRVHALVDAFQSRSSSVL 361
OY 336 HKTKNIERMKHLKTIEMVPTFKTOSPCGVFSKLGVLRRV-----TRNLVKNKLAVI 387
Db 362 YATPLSMKEETNGMRPRKKAIVERTDGMWROFELLKRAMMOASRDGPTNKNVRRARMSVA 421
OY 388 TRLQNLIMGLFLEFVLNRSNVLKGAIQDRVGLTYQVGAFTPYTGMLNAVILFPLRA 447
Db 422 SA-----VIGSVFWRMGKSGTSTIODRMGLQVAINTMAALTKTVGVFPKERA 471
OY 448 VSDQESQDGLYQKWMMLAYALHVLPEFSVAATMIESSVCYWTGLHPEVARFGYFSAALL 507
Db 472 IYDRERSKGSYSLGPTLLSKTIAEIPIGAAPLMFGAVLYPMARLNPPLTSRFGKFCGIY 531
OY 508 APHLIGEFLTVLIGIVONPNVNSVALLSTAGVYVSGFELRNIOEMPIPKIISYTFE 567
Db 532 VESFAASAMGLTVGAHAVPSTEAAMAVGPSLMTVFYVFG-GYYVNDNPTPIFRWIPRASI 590
OY 568 QRYCEILVYNEEYGLNF 585
Db 591 IRMAFOGLCINFEGLKF 608
```

Search completed: June 11, 2003, 09:13:23
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:09:57 ; Search time 22 seconds
(Without alignments)
2844.705 Million cell updates/sec

File: us-09-989-981a-6
Perfect score: 3326
Sequence: 1 MCDLSSVTPGSGMGLAVNRG.....PALVILGIVFKIRDLISR 651

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649.5	19.5	725	2 T47652	ABC transporter-11
2	627	18.9	1294	2 S77690	probable membrane
3	621	18.7	1049	1 S19421	ATP-dependent perm
4	614	18.5	720	2 T47648	ABC transporter-11
5	610.5	18.4	725	2 C84423	probable ABC trans
6	608	18.3	708	2 T47650	ABC transporter-11
7	602.5	18.1	687	1 FYEPM	white protein - fr
8	597	17.9	590	2 B96573	protein F12M6.17
9	595.5	17.9	658	2 T31958	hypothetical prote
10	591.5	17.8	740	1 T02567	probable ATP-bindi
11	590.5	17.8	638	2 G02068	white homololog - hu
12	589.5	17.7	646	2 C86441	probable ABC trans
13	584	17.6	755	2 G84791	probable ABC trans
14	583.5	17.5	687	2 D96553	hypothetical prote
15	578.5	17.4	646	2 JC7777	ATP binding caset
16	575.5	17.3	649	2 A84509	probable ABC trans
17	572	17.2	739	2 T45891	ABC transporter-11
18	571.5	17.2	609	2 E96742	probable ABC trans
19	564	17.0	662	2 T47649	ABC transporter-11
20	554	16.7	559	2 B88474	protein C05D10.3
21	552.5	16.6	635	2 T08934	hypothetical prote
22	545.5	16.4	608	2 T34391	hypothetical prote
23	540	16.2	678	2 H96552	hypothetical prote
24	539	16.2	577	2 T04429	ABC-type transport
25	529.5	15.9	633	2 T19189	hypothetical prote
26	527.5	15.9	659	2 E86313	hypothetical prote
27	517.5	15.6	610	2 T19333	hypothetical prote
28	517.5	15.6	639	2 G88839	protein C10C6.5 [1
29	514.5	15.5	705	2 D84680	probable ABC trans

30	504	15.2	695	2 T21109	hypothetical prote
31	497	14.9	547	2 T31543	hypothetical prote
32	488	14.7	737	2 T46101	ABC transporter-11
33	464	14.0	1450	2 A84780	probable ABC trans
34	456.5	13.7	1426	2 T30567	ATP-binding caset
35	454.5	13.7	675	1 FYEPM	brown protein - fr
36	435	13.1	668	2 S55023	brown protein - fr
37	429	12.9	1469	2 H96522	probable ABC trans
38	425.5	12.8	1443	2 T02451	probable ABC trans
39	425	12.8	1451	2 B86286	F9L1.15 protein -
40	421	12.7	1450	2 T45888	ABC transporter-11
41	420	12.6	1435	2 D96593	protein putative A
42	418.5	12.6	1420	2 T02644	ABC-type transport
43	417.5	12.6	1619	2 T30541	ABC1 transport pro
44	413	12.4	1413	2 G84790	probable ABC trans
45	410	12.3	1177	1 D71416	probable PDR5-like

ALIGNMENTS

RESULT 1					
T47652					
ABC transporter-like protein - Arabidopsis thaliana					
N:Alternate names: protein T26112.10					
C:Species: Arabidopsis thaliana (mouse-ear cress)					
C>Date: 20-Apr-2000 #sequence, revision 20-Apr-2000 #text, change 19-May-2000					
C:Accession: T47652					
R:Monfort, A.; Casacuberta, E.; Pulgomech, P.; Mewes, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, February 2000					
A:Reference number: Z24471					
A:Accession: T47652					
A>Status: preliminary					
A:Molecule type: DNA					
A:Residues: 1-725 <MON>					
A:Cross-references: EMBL:AL132954					
A:Experimental source: cultivar Columbia; BAC clone T26112					
C:Genetics:					
A:Map position: 3					
A:Note: T26112.10					
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12M6.1;					
Query Match					
Best Local Similarity 29.4%; Pred. No. 1e-40; Length 725;					
Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;					
QY	33	PEPSIGILHASYSVSHRVRPMDITSCROMPROIIDVSLVSGQICIGSSGSK	92		
DB	68	PVYVYLNFMNLOVDVITARR-----FGFSRONGVKTLLDDVSGEASGDILAVLAGSAGK	123		
QY	93	TTLLDAMSGRLAGAGFLGEVYNG-RALREDFQDFSVYASDITLLSLVRETLAYT	151		
DB	124	STLLDALAGVAE-GSLRGSVTLNGEYVLSRLKVISAVYMDDLFPMVYKELTMA	182		
QY	152	ALLAIRRG-NPGRQKVEVMAELSHVADRLIGVSLGSGTSGRRVSTAAQLD	210		
DB	183	SEERLPISLSKSKMEVEVALIDQLGRNANAVIVGEGHGVSGGRRRVSIGIDITHD	242		
QY	211	PKVMEDEPTGLDCMTANOIVLVLEARRNRIVVTHOPSELEFOLDKIAISFGE	270		
DB	243	PVLFDEPTSGIDSTNAFVVOYVKRIAGSGSVISHQPSARIYELDRILLISRGK	302		
QY	271	LIFCGPAENLDEFNDGYPCEPSHNFDEYMDLTSVDTOSKREIETS-----	319		
DB	303	SVENGSPASIPGCFSDGRPIPEKENTISERALDV-----RELGSNEGKALVDEN	354		
QY	320	-----KRVOMEESAYK-----KSAIGHKTL--KNIRMAHLKTLPAVPRPTKD	360		
DB	355	EKWQOKNISLQSPQTNKLDORSLSLKAINASVSGRLVSGSSNSNPTSMETVSSYA	414		
QY	361	SPGVFSKLGVLRRVTRNLVTRNLAVITRLQNLINGLFLFFVLRVRSVYVKGADQRY	420		
DB	415	NPSLFETF-ILARRYKKNWIRMPDELVGTIRIATVAVTGC-TLATVYWKLDHTPRGA-OERL	471		

```

QY      421  GLLYGVGAPPTGCMNANVLPVLRANVSDQESDGLQKQMMPLAALAHLPSPSVATM 460
Db      472  -TLFAVVPVPMFCCLDNVPVFIQBRVIFLRETHNAVRTSSYVISHSLVLPDLPASP 530
QY      481  IFSSVCWTGLGHPPEARFGYPSAALLAPHLIGEFVLVLGFIQONPIVMS--VVALTSI 539
Db      531  VFSAINFVGVLSGGLEGSEFVECLLIYASFPMSGSSVYFISGV--FRIMCIWVSTIYL 568
QY      540  AGVLVSGGFLRNIOEAPIDPKIIISYFPQKYCSEILVYNEFYGLNFGSGSNVSVYTNPM 599
Db      589  AYCILLSGEYVNDRIPIFYWTWPHYIISILKPYEAVLINE-----DDPS 633
QY      600  CAFTQGIQIFLEKTCPGATS 618
Db      634  KCFVRGVQVDFSTLLGVS 652

```

RESULT 2

probable membrane protein YOL075C - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein 01125, hypothetical protein 01130, hypothetical
C:Species: *Saccharomyces cerevisiae*
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
C:Accession: S77690; S66767; S66768
R:Alexandrakaki, D.; Katsoulou, C.; Tzermia, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66756
A:Accession: S77690
A:Molecule type: DNA
A:Residues: 1-1294 <ALE>
A:Cross-references: EMBL:Z74816; MIPS:YOL075C
A:Note: this is a revision to the sequence from reference S66756
A:Accession: S66767
A:Molecule type: DNA
A:Residues: 1-179, 'TTRTGVFLVKKRED' <ALM>
A:Cross-references: EMBL:Z74816
A:Experimental source: strain S288C
A:Note: this sequence has been revised in reference S77690
A:Note: this was assumed to be protein YOL074C
A:Accession: S66768
A:Molecule type: DNA
A:Residues: 200-1294 <ALP>
A:Cross-references: EMBL:Z74817
A:Experimental source: strain S288C
A:Note: this sequence has been revised in reference S77690
A:Note: this was assumed to be the complete sequence of protein YOL075C
C:Genetics:
A:Cross-references: SGD:S0005435
A:Map position: 15L

C:Superfamily:Unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP, nucleotide binding, P-loop, transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:376-392/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:506-622/Domain: transmembrane #status predicted <TM4>
 F:710-916/Domain: ATP-binding cassette homology <ABC>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1042-1058/Domain: transmembrane #status predicted <TM5>
 F:1125-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match

Query Match 18.9%; Score 627; DB 2; Length 1294;
Best Local Similarity 31.7%; Pred. No. 1.1e-38;
Matches 181; Conservative 106; Mismatches 228;
Indels 56; Gaps 19;

Dy 65 TROILKVDLSLYESGOIMC ILGSSSGSKTTL LDANGRLGRAGTFEL----GEVYNGRA 119
|::||: |: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 706 TREILQSVNAIFRPGMINAIMGPSGSKSLNLNLSGR L-KSSVFARFDTSISIMENDIQ 764

[illegible]

RESULT 3

ATP-dependent leucinease ADP1 precursor yeast
N:Alternate names: protein YCR011c; protein YCR105
C:Species: Saccharomyces cerevisiae
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: S19421; S40914
R:Goffeau, A.; Purnelle, B.; Skala, J.
submitted to the Protein Sequence Database, March 1992
A:Reference number: S19420
A:Accession: S19421
A:Molecule type: DNA
A:Residues: 1-1049 <GOF>
A:Cross-references: EMBL:X59720; MID:g1907116; PIDN:CAA42328.1; PID:g1907154; GSPDB:
R:Purnelle, B.; Skala, J.; Goffeau, A.
Yeast 7, 867-872, 1991
A:Title: The product of the YCR105 gene located on the chromosome III from Saccharom
A:Reference number: S40914; MUID:92160395; PMID:1789009
A:Accession: S40914
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1049 <PUR>
R:Skala, J.; Purnelle, B.; Goffeau, A.
Yeast 8, 409-417, 1992
A:Title: The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm
K:genes
A:Reference number: S25353; MUID:92327849; PMID:1626432
A:Contents: annotation
A:Genetics:
C:Gene: SGD:ADP1; MIPS:YCR011c
A:Cross-references: SGD:S0000604; MIPS:YCR011c
A:Map position: 3R
C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
;1-25/Domain: signal sequence #status predicted <SIG>

F:26-1049/Product: ATP-dependent permease ADP1 #status predicted <MAT>
F:26-324/Domain: extracellular #status predicted <EXT>
F:323-341/Domain: transmembrane #status predicted <TM1>
F:406-607/Domain: ATP-binding cassette homology <ABC>
F:423-430/Region: nucleotide-binding motif A (P-loop)
F:550-557/Region: nucleotide-binding motif B
F:794-810/Domain: transmembrane #status predicted <TM2>
F:829-845/Domain: transmembrane #status predicted <TM3>
F:878-894/Domain: transmembrane #status predicted <TM4>
F:909-925/Domain: transmembrane #status predicted <TM5>
F:938-954/Domain: transmembrane #status predicted <TM6>
F:1055-1041/Domain: transmembrane #status predicted <TM7>
F:50,114,165,221/Binding site: carboxylate (Asn) (covalent) #status predicted
F:42/Binding site: ATP (Lys) #status predicted

enus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-725 <STO>

A:Cross-references: GB:AE002093; NID:94262239; PIDN:AAD14532.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g01320

A:Map position: 2

Query Match

Best Local Similarity 29.7%; Score 610.5; DB 2; Length 725;
Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

```

QY 51 VRP-----WMDITSC-----RQMTROLKQVSLVESGQIMCIGSSGSKTLLDMSGR 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 IRPVTIRMNITCSLSDSKSVRFLLKNVSGEAKGRLAIWGPSSGKTTLLNVIAGQ 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 IGRAGT--FLGEYVNGRALLRREPQDCSYVLSQSTLLSLVRETTALTALARR-G 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 LSLSPRLHSLGLEPVGKSSSKAYK--LAFVRQEDLFFSOLTVRETTLSFAELQIPETS 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 NGSFOKKYEAIVMAELSLSHVADRLLIGNSLGISTGERRRYSIAQLIDPKVMLFDEP 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 183 SAEERDEYVNNLLKLGIVASCADSCVGDAKVNGISGEEKRLSLACELIASPSVIFADEP 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 TTGLDCKMTANQIVLVLELARRNRIVYLTIHOPRSELPOLFKIALISFGLIFCGTPA- 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 243 TTGLDAFOAEKVMETLQKLAODGHVYICSIHQPRGSVYAKFDIYLLFEGTIVYAG-PAG 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 -EMLDFNDGYPCEPHSNPFEDYMDLTSVDTSQSKRELETESKRVOMIESAY--KSAIC 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 302 KEPLTYTFNGFGLCEPHVPAEFLADLVSDYSSSEYVSSQRRHALDASQSSSVL 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 336 HKTINIERMKHLKTLPMVPFTKDSPGVFSKLGVLRLV-----TRNLVRNKLAVI 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 YATPLSMKEETKNGMPRRKAIYERTDGMWRQPFLLKRAMQASNDQTNVBARMSVA 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 TLLQNLIMGLFLFLFVLRVRSNYLKGAIQDRAVGLLYQVGAFTPTGMLNANVLEPVLDA 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 422 SA-----VIGSYVMRMKGSQTSIQDMMGLIQAINTAAMALTKYGVPEKERA 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 448 VSDQSDGLYOKMOMLALYALHVPFSVATMIFSSVYWTGLHPEVARRGYSAALL 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 472 IYDRERSKGSYSLGPIILSKTIAETIPICAFPLMGAVLYPMARLNPILSRGKFCGIVT 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 508 APHLIGELTLLVLLGIQONPNIVNSVVALLSIAGVLSGFLRNIOEMPIPKIISYTFE 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 533 VESFASAMGLTVGAMVPSTEAMAVGPELMTVFIVFG-GYVYVNDNPIIFRWIIPRASL 590
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 566 OKYCEILLVNEFYGLNF 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 591 IRMAFOGLCTINEFSGLKE 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6

T47650
ABC transporter-like protein - *Arabidopsis thaliana*
W:Alternate names: protein T15C9.110
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence-revision 20-Apr-2000 #text-change 19-May-2000
C:Accession: T47650
R:Reviews, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47650
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <MEM>
A:Cross-references: EMBL:AL132970

A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.110
C:Superfamily: *Arabidopsis thaliana* probable ATP-binding cassette protein P12L6.1;

Query Match

Best Local Similarity 27.5%; Score 608; DB 2; Length 708;
Matches 111; Conservative 136; Mismatches 256; Indels 58; Gaps 15;

```

QY 18 NNGSGSLGAP--TAPPHSLGILHASYSVSHRVPRMWDYSCROMTROLKDVSLY 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 41 NAPTQHILDLAPAEATRSVPFLSFNNLSYNNVLR--FDSRRRTAATVKTLLDDITSE 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 VESGQIMCIGSSGSKTLLDMSGRLLRAGTFLCEYVNG-RALRRQFDDCESTVIG 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 99 ARDGETLAVIGSGGAKSTYLDALGRVAE--SLKGTVNLNGEKVQLSKLKVISAIVAQ 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 SDTLLSLVRETLHTALTALIRGPGSFO--KRYVAWAEISLHVADRLLIGNYSLGIT 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 158 DDLPPMLTVKETLFASEFRLPRSLPKSKMERVETLIDQLRNAADTVIGDEGRGV 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 194 STGERRRYSIAQLIDPKVMLEDEPTGIDCKMTANQIVLVLELARRNRIVYLTIHOPR 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 218 SGGERRRYSIGIDIHDPILFLDEPTSGLDSTNFMVVOYLKRIQSGSVYIMSHOPS 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 254 SELPOLFKIALISGELIFCGTPAEMLDFFNDGYPCEPHSNPFEDYMDL-----T5V 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 278 ARIGLIDRLILSHKSVFNGSPVSLPSFSSFGPRPEKENITEFALDVIARELGSSSE 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 308 DT-----OSKERIEFSKRVOME-----SAVKSATCHTKIEMERKHLKTLIP 352
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 338 GTRDLVEPRKQONQATATQOSRSLKALIAVSRKLLVSGSGAGAPISMETVSYA 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 353 MVEPFTKDSPGVFSKLGVLRLVNRNLYRNKLAVYTRRLQNLIMGLFLFLFVLRVSNYL 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 NPP-----LAEFLIAKRYIKMWTPELIGRIGCTVWGTGLLATVYRRL-DNRP 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 413 KGAIDRVCGLYQFQGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKMOMLALYALHVL 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 448 RGA-OERMG-PFAFGSTVFYCCADNIPFQIDERYIFLETHNMYRSTSYISHALVSL 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 PFSVATMIFSSVYWTGLHPEVARRGYSAALLAPHLIGELTLLVLLGIQONPNIVNS 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 506 PQLMALSTAFATTTWVGLSGLESFYICLIITAAFTSGSIYTFISGL--PVMMS 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 533 -VVALLSIAGVLSGFLRNIOEMPIPKIISYTFPOKYSILVNEFYGLNFTGSSN 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 564 YWTVIAYLSYCLLGGFYINRDIRIPLYMIFHYISLKKPYEAVILNER----- 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 592 VSVTNPMACTOGIOFIETK 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 613 ----DPSRCFVGQVFDGT 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

FFFW

white protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 31-Dec-1990 #sequence-revision 17-Feb-1995 #text-change 19-Jan-2001

C:Accession: S08635; S07263; S10240

R:Pepling, M.; Mount, S.M.

Nucleic Acids Res. 19, 1633, 1990

A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A:Reference number: S08635; MUID:90221897; PMID:109311

A:Accession: S08635

A:Molecule type: mRNA

A:Residues: 1-687 <PEP>

A:Cross-references: EMBL:X51749; NID:98825; PIDN:CA36038.1; PID:98826

R:O'Hare, K.; Murphy, C.; Lewis, R.; Rudin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A:Reference number: S07263; MUID:85134865; PMID:6084717

A:Accession: S07263
 A:Molecule type: DNA
 A:Residues: 1-24, 'LIFEIPYHCRVAD', 30-334, 'ITLHLSYAWVPSVLPITIRRTTYRCNPLCPDGSSPPTGSPRI
 A:Cross-references: EMBL:X02974
 A:Experimental source: strain Canton S
 R.O'Hare, K.
 Submitted to the EMBL Data Library, June 1985
 A:Reference number: S10240
 A:Accession: S10240
 A:Molecule type: DNA
 A:Residues: 1-24, 'LIFEIPYHCRVAD', 30-687 <OH42>
 A:Cross-references: EMBL:X02974; MID:g10873; PIDN:CAA26716.1; PID:g10874
 A:Experimental source: strain Canton S
 C:Genetics:
 A:Gene: white; w
 A:Cross-references: Flybase:Fbgn0003996
 A:Introns: 24/3; 116/1; 334/2; 439/3; 483/3
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
 F:113-317/Domain: ATP-binding cassette homology <ABC>
 F:130-137/Region: nucleotide-binding motif A (P-loop)
 F:261-265/Region: nucleotide-binding motif B
 F:67,93,472,554,651/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.1%; Score 602.5; DB 1; Length 687;
 Best local similarity 28.8%; Pred. No. 3.2e-37;
 Matches 180; Conservative 131; Mismatches 220; Indels 95; Gaps 19;

QY 66 ROLIKVSLYVESGQIMCIGSSGSGKTLLDAMSGR--LGRAGTFLGEYVNGRALRRE 123
 DB 110 KILKNCVAYPGDGLLAVWSSGAGKTTLLMLAFPSPGICQVSPSGMLNLQPDVAK 169
 QY 124 OFODCSYVLSPTLLSLVRETIHTALTRRNGPSGOK--KVEAVMELSLSHV 180
 DB 170 EMORCAVYQODLFLGSLTAREHLIFQAN--VAMPRLTYRKQVAVDVIOELSLSKC 227
 QY 181 ADRLIG--NYSIGISGTERRRVSIAMQLDPPKMLFDEPTTGLDCMTANQIVLVLELA 239
 DB 228 QHTIGVPGKGLSGGERRLAFASALTDPLLLICDEPTSGIDSTAINSVQVAKLS 287
 QY 240 RRRRIYVLTTHORSELPOLFKIALISFCELIJFCGPALMDLDFNDCGYPCEHSPFD 299
 DB 288 QOKRTYILTHOSSEIFELFKILLMAEGRAVAFIPPSBAVDFFSYVGAQCPYNPAD 347
 QY 300 FYMDLTSVDSQSERELSKRYOMIESAKKSAICKTKTKNIENRKHLLTLPVPPKTK 359
 DB 348 FTVQVLAIV--VGRLEIESDRITAKICDNFAIS--KVARDMQELATATNLE--KPL 396
 QY 360 DSP-----GVFSKIGVLRRTVRLVNRKLAIVTRTLNGLFL--LFFVLRYSN 410
 DB 397 EDPENGYTKATFMOPRAVLAMRSMLSVLKEPLLVKRLQTTMVAIILGILFLGQQLTQ 456
 QY 411 VLKGAIQDRVGLLYQFVGATPYTGMLNANVLFVPLRAVSDQSDGLYQKQMMALYALH 470
 DB 457 V--GWMNNGAIFLELTJMTFQNFATINVFSELVPMREARSLRYRCDYFELGRTIA 513
 QY 471 VLPFSVATMIPSSVCCYTWLGLHPEVARGCYSAALLAPHLIGEPFLTVLLGLIVONNIV 530
 DB 514 ELPLFLTVPLVFTALAYPMIGLRAGLVH-----FNCALAVTVV--ANVS 556
 QY 531 NSVALLSIAG-----VLVSGFLNIQEMPIPEFKIISYTFQKXCEI 574
 DB 557 TSEGYLISCASSSTMAISVGPVITLPELFGFELNCGSVPYLMLWLSLFRANEG 616
 QY 575 LVVNEEYGL--NFTGSSNVSVTTPMCAFTOGIOFIKTCP--GATSRFTNNFLIYS 629
 DB 617 LTINGADVEPGEISCTSSNT-----TQSSSKVILETILNFA--A 655
 QY 630 FIP-----ALVILGIYVFKIRDLISR 651
 DB 656 DLPDYGLAIV--IVSEFRLAYLALR 680

RESULT 8
 Protein F12M16.17 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: B96573
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Jansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Jansen, N.F.; Hughes, B.; Hulzar, L.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Mai
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tr
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: B96573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-590 <STO>
 A:Cross-references: GB:AE05173; MID:g769856; PIDN:AAF69534.1; GSPDB:GN0141
 C:Genetics:
 A:Gene: F12M16.17
 A:Map position: 1
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.9%; Score 597; DB 2; Length 590;
 Best local similarity 29.6%; Pred. No. 6.7e-37;
 Matches 186; Conservative 113; Mismatches 270; Indels 60; Gaps 15;

QY 29 PATRPE-----HSLGILNARSYSRVRPMDITS--CRQOTROLKDVSLYVESGQI 81
 DB 4 PVKAPIGGRIISRTETKMLSYRIGNTKRFNSMCLSEKEKVKILKVSQCARSAEI 63
 QY 82 MCILGSSGSKTTLTLDAMSGRLGRAGTFLGEYVYVNGRALRQODCSYVLSPTLLSS 141
 DB 64 TALNPGGAGKTTILLETILAGVSH--GVYSQVLYNPGMPPEYRRVSGFVPODALPPE 122
 QY 142 LVVRETIHTALAIRRNGNSFPQKVEAVMAELSLSHVADRILGNSYLSGISTGERRV 201
 DB 123 LTVQETLTSALRLKTKRDA--AAKRLIQLGELHVAADSRIGSSRSGISGERRV 181
 QY 202 SIAQLQDPKVMLEFDEPTTGLDCMTANQIVLVLELA--RRRIYVLTTHORSELPOLF 260
 DB 182 SIGELVHDPNVILIDEPSTGLDSASALQVTVTLKMTIKQKTYVLTTHQPFRILEQI 241
 QY 261 DKIALISFGEILFCGTPAEMLDFNDCGYPCEHSPFDYMD-----LTSVDNQSKREI 316
 DB 242 DRYVLNSNGVONGSYSLHOKIKESGHOIPRRVNLLEYAIDIASLEPIRTQSC--REI 300
 QY 317 ETSKRYOMIESAVKSAICKTKTKNIENRKHLLTLPVPPKTKDSPCVFSKIGLRLRV 376
 DB 301 -----SCYGH-----KTKMSC-----YISAGGLHSDSHSNVLEVOITLGRSC 342
 QY 377 RNLVRKLAIVTRLLNGLFLLFEVLNRSNVLKGAIQDRVGLLYQFVGATPYTGML 436
 DB 343 KNIFRTKQLFTTRALQISAGLLIGSYLVN--GNOKKEAVLRFG--FEAFILFLLSST 400
 QY 437 NAVNLPFVLAVDQSDGLYQKQMMALYALHLPFSVATMIPSSVCCYTWLGLHPEV 496
 DB 401 EGPLFLQDRILIRRETSRAVYLSVLAIDLIFELFLIIMELFATPYWVGLRREL 460
 QY 497 ARGYSAALLAPHLIGEPFLTVLLGLIVONNIVNSVALLSIAGVLYVSGFLRNIQEM 556
 DB 461 DGLFYSVIVITVILKNSNVACFSALVPEFINGTIVISGL--NGSFLLSGFYIANDRIP 519
 QY 557 IPPKIISYTFQYCCSEILVNEFGLNFTGSSNVSVTTPMCAFTOGIOFIKTCPCGA 616
 DB 520 VVVEFVHVLSTFKYPECLMINEX-----RDGVFLKQDDLE 556
 QY 617 TSEFTMNLIFSFIATVILGIYVFKIR 645

QY 347 ----HLKTLPMVPEFKTKDSQVFSKGLVLLRVRTNLRVKNKLAVITRLLONLIMGLEFLF 402
 DB 415 GATTNTTTLAVPAFPAP-----MWIEIKTLKSRMLNSRQPELFGIRIAGVITG-FILA 469
 QY 403 FVLAVRSNVLKGAIDORVGLLYQFGATPYTGMLNVLNFPVLRANSDOSSQGLQOKMO 462
 DB 470 TVEFWRLDNSPKG-VQERLIG-FFAFAMSTMEYTCADLIPVLOERYITFMRATVNAVARRSS 527
 QY 463 MMLAVLHVLPESVAVTMFFSSVCWTGLGLPEVARGPSAALLPHLIGFEFLTVLLG 522
 DB 528 YVLSHAIYSPSLIFLSVAFAATTYNAVGLDGLTGLIFCLITLISFMSGSSFFVFLSG 587
 QY 523 IVONPNIV---NSVALLSTAGLVGSGFLRNIOEMPIPEKIIISYTFPOKCEIIVVNE 579
 DB 588 VV--PSVMLGTYTIVAIL--AYPLFSGFFINRNIPDVMWFMHMSLVKYPYEVNLONE 643
 QY 580 FYGLN--FTCG-----SSNVSVTTPNMCATQO----- 605
 DB 644 FSDATKCFVAGVQIFDNTPLGELPEVMKLLGLTVKRSIGVITSTTCITFGSDILRQOG 703
 QY 606 -IOFIETCPGATSRFTMNFLLYSFIPALVIIG 638
 DB 704 VVQLSKNCLFTVAGGFFRILIFY---TLILG 734

RESULT 11

G02068
 White homolog - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001
 C:Accession: G02068
 R:Crop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Archilegas,
 submitted to the EMBL Data Library, August 1995
 A:Reference number: H00769
 A:Accession: G02068
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <CRO>
 A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277
 C:Genetics:
 A:Gene: white
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop
 F:61-253/Domain: ATP-binding cassette homology <ABC>
 F:78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.8%; Score 590.5; DB 2; Length 638;
 Best Local Similarity 26.3%; Pred. No. 2.3e-36;
 Matches 164; Conservative 142; Mismatches 266; Indels 51; Gaps 14;
 QY 44 SYSVSHRVPMMDITSCROQWROLIKDVSIVYESQIMCIGSSSGKTTLLDMSGR 103
 DB 43 STSVPE--GPMW-----RKKGKTKLKGISGRFNSGELVAVIMGPSGAGSTLMNIIAGY- 94
 QY 104 GRAGTFLGFEVYNGRALRREOFQDFCFYVLOSITLLSLVRETELHATLLAIRGNPSS 163
 DB 95 -RETQMGKAVLINGLPDRLCFRKRYSCYIMQDKMLPHITVQEDAMVSAHLKIQEODEGR 153
 QY 164 FOKKAYVAEELSHVADRLIGNSLGISTGERRRYSIAAQLDOPVLMFDEPTTGL 223
 DB 154 -REMKETILTALGLSCA-----NTRTGSLSGQKRRLAILELVNPNVPMFDEPTTGL 207
 QY 224 DCMTANOIVLVLELARRRIVVLTTHORSELFOLEFKIALISRGELIFCGTPAMDM 283
 DB 208 DSASCPQVAVSLMKGLAOGGRSTICTHQSARKLAFELFDLYVLSOGQCYRGRKVCNLVY 267
 QY 284 FNDGCPPEHNSPFDYMDLTSDVTPQSKEREIETSKRVOMTESAVKKA-----I 334
 DB 268 LRDGLINCPYHNPADEFVAVEASGEIGDQNSRLYAVRGMCDSDHKRLGDAEYVNP 327
 QY 335 CKTKTKNTERMKHLTKLTPVAPKTKDSQV-----FSKGLVLLRVRTNLRVKNKL 384

DB 328 WHRSEEVQKTKRLKL-----RKDSSMEGCHSFSSASCTFOCLIFKRRFLISIMRDSV 381
 QY 385 AVITRLLQNLIMGLEFLFVLRVRSNVLKGAIDORVGLLYQFGATPYTGMLNVLNFPV 444
 DB 382 LTHLRITSHIGILIGLILGLYLGIGNEAKK--VLSNSSEFLFSMLFELFALMPVLEPL 439
 QY 445 LRAVSDQSDQGLYOKOMMAYALVHLPESVAVTMFFSSVCWTGLGLPEVARGPSA 504
 DB 440 EMGFTLEHNLNMYSLKAYLLATKMTADVPOIMEFPVAVCYVWMTSQSPDNAFVFAA 499
 QY 505 ALAPHLIGFEFLTVLLGIVQNPVYNSVALLSIAGLVGSGFLRNIOEMPIPEKIIISY 564
 DB 500 LGTMTSLVAVOSLCL-LTGAASTSIQVATFEVGPVTAIPVLLFSGFEVSEFDITPYLQWMSY 558
 QY 565 FTQOKCEIIVVNEFYGLNFTPGSSSVSTTPNMCATQOIFETTCGATSRFTMNF 624
 DB 559 ISTVRIGFEVGLIS-ITGLD---REDLCHDIDETCHF-OKSEALIELDVENAKYLD 612
 QY 625 LIYSEFIPALVIIGIV--FKIR 645
 DB 613 IVIGIFPISRLIATVPLRYKIR 635

RESULT 12

C86441
 Probable ABC transporter [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86441
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; A
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Deva
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; K
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maitl, R.; Mar
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Ta
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; M01D:21016719; PMID:11130712
 A:Accession: C86441
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-646 <SRO>
 A:Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

Query Match 17.7%; Score 589.5; DB 2; Length 646;
 Best Local Similarity 29.9%; Pred. No. 2.8e-36;
 Matches 183; Conservative 112; Mismatches 246; Indels 71; Gaps 22;
 QY 10 GGSM--GLAVNRSGS--SLGAPATAPPEPISLITLASVS-----HRYRPMMDITSCR 61
 DB 14 GGVNAGGLPDMSTQSKSVIATPITISQ---GLQSMPTITLKEVYKVK-LEQTSQCA 69
 QY 62 QGV---TROLKDVSLYESGOIMCIGSSSGKTTLLDMSGRLAGTFLGFEVYNGR 118
 DB 70 GSWKSKERTLINCITGMVCPGEFLAMGPSGSKTLLSALGRLSK--TFSGKYVNG- 126
 QY 119 ALRREFOQCF-----SYVLOSITLLSLVRETELHATLLAIRGNPSPQK-----KYE 169
 DB 127 ----QPSGCCIRRRGFEVADVDVLTPLTWETLFTALRLT-----PSSLTRDEKAEND 178
 QY 170 AVAAEELSHVADRLIGNSLGISTGERRRYSIAAQLDOPVLMFDEPTTGLDQMTAN 229
 DB 179 RVIAELGLNRCINSMIGLGLFRGISGEEKRRYSIGQEMLINSLLADEPTSGLDSTYAH 238
 QY 230 QIVVLLVLELARRRIVVLTTHORSELFOLEFKIALISRGELIFCGTPAMDMFNDGCV 289
 DB 239 RIYTTIKRLASGGKRVVTTTHOPSSRIYHMPKQVLLSGSPITTYGAASAVYEYFSSLGF 298


```

Db 371 -----RELSTPSTGTPLEVEFHKQWAKAPSTNNN---NKRNTNWSLKEAITASTSRGK 423
QY 333 -----MVD-FRTKDSPGVYSKGLVLLRRVTRNLVNRKLAVITRLLQMLTJGLE 399
Db 424 LVSGATNNSSNMLTFESFQTFANP-FWIEIMIVIGKRALILNRSRQPELLGRLGAVMTGTII 482
QY 400 LLEFLVRYRNVLKEAIDRWGLLQFQNGATFYTGMLNVNLEPVLRAVSDQSGCLQ 459
Db 483 LATMTNL-DNSPKKA-ORLQ-FYAFANSTIFYCAEALPVFLORRYTFMRETAANAAR 539
QY 460 KWQMLAYALHVLPEFSVATMIFSSVCYWTGLQIHEVAFGFSALLAPHLIGFEFLTY 519
Db 540 RSSYLSQSIISIPALIVLASFAATTFEAVGLDGGANGEFFEYFIILASFNAGSSFYTF 599
QY 520 LIGTIOGNIV--KSVYALLSIAGVLVSGGSLRNIQEMPIFFKIISTYFOKYCEILV 576
Db 600 LSGV--PNVMIGFTVVAAIL--ATFLFSSGFIISDRIPYMLFMHYISLVKYPIREGVL 655
QY 577 VNEFYGLNFTGSSNVSVTYTNPMCAFTOGIOGIEKT 612
Db 656 ONEF-----QNTRCFAFGVOLFDS 676

```

RESULT 14
D96553
hypothesis

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D96553
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Allen, N.F.; Hughes, B.; Huiter, L.
C:Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kitz
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A6141; MUID:21016719; PMID:11130712
A:Accession: D96553
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005173; NID:g10092349; PIDN:AA612758.1; GSPDB:GN00141
C:Genetics:
A:Gene: F5D21.6
A:Map position: 1
Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; 2

Query Match	17.5%	Score 563.5	DB 2	Length 687
Best Local Similarity	28.0%	Pred. No. 8	6e-36	
Matches	167	Conservative	117	Mismatches 239; Indels 73; Gaps 16;
QY	33	PEPSLGLIHASYSVSHRPRMMDITSCROOQ----	TRQILKDVLSYVSGQIMCIGSS	88
		: :	:	: : :
Db	13	PPPAEIG--RCAYIA-----	WEDLTIVIPNMSGGPTRLRLDGLNGHAPGRIMALMGPS	64
QY	89	GSGCTTLLDAMSGLRAGTFLGEYVYVNGRALRPFQDCFEYVLQSDTLLSSLYTRETLL		148
		: : :	:	: : :
Db	65	GSGSTLLDSTLAGRLARVINTGILLNGKARLD--	YGLVAYVYQEDITLMTGLTYRETI	122
QY	149	HYTALLAIRNGNPSPSFOK-----	VEAWAELTSLSHVADRILGNTSLGISTGERRRVSI	203
		: :	: : : :	: : :
Db	123	TYSAHLRL-----	SSDLKREEVNDIEGTHIELGLQDCADRYIGNHSHKRVSGGERRRVSI	178
QY	204	AAQLLDPKYMLPDEPTTGLDCMTANQIVLVLLVEAR--RNRIVLVLIHQPNSELFQLEDK		262
		: : : :	:	: : :
Db	179	ALFELIRPQILLFDEPTSGLDASAFVYQLALRNIRADRGRTVYVSIHQPSSEVFALFEDD		238
QY	263	IALISFQELFFCGTAPMLDFFNDGCGVCPCEPSHNPDPFYMDVTSVDTSKREKIEFESKRV		322
		: : : :	:	: : :
	239	LEFLSSSEYVYFGSGKFAVEFFAFAGGKPKPKRRSDHFLACTNSDPTTVAATVIGSOTI		299

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:12:22 ; Search time 25 Seconds

(without alignments)
2688.377 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MGDLSTLRFGSGMGLOVNRG.....PALVILGIVFKIRDLISR 651

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US08_NEM_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubppa/PCT_NEM_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppa/US06_NEM_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppa/US07_NEM_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppa/US09_NEM_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppa/US10_NEM_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppa/US60_NEM_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3326	100.0	651	US-10-090-455-6	Sequence 6, Appl1
2	3326	100.0	651	US-09-989-981A-6	Sequence 6, Appl1
3	3326	100.0	651	US-09-837-992-3	Sequence 3, Appl1
4	2744.5	82.5	652	US-09-989-981A-2	Sequence 1, Appl1
5	2744.5	82.5	652	US-09-837-992-1	Sequence 1, Appl1
6	697	21.0	672	US-09-989-981A-4	Sequence 4, Appl1
7	697	21.0	673	US-10-090-455-7	Sequence 7, Appl1
8	697	21.0	673	US-09-989-981A-8	Sequence 8, Appl1
9	682.5	20.5	655	US-09-961-086-1	Sequence 1, Appl1
10	680.5	20.5	655	US-09-981-353-35	Sequence 35, Appl1
11	680.5	20.5	655	US-10-120-687-61	Sequence 61, Appl1
12	674.5	20.3	655	US-10-090-455-5	Sequence 5, Appl1
13	674.5	20.3	655	US-09-866-866A-10	Sequence 10, Appl1
14	672.5	20.2	655	US-09-866-866A-27	Sequence 27, Appl1
15	660	19.8	657	US-09-866-866A-14	Sequence 14, Appl1
16	602.5	18.1	664	US-10-108-605-245	Sequence 245, App
17	598.5	18.0	674	US-10-090-455-4	Sequence 4, Appl1
18	590.5	17.8	638	US-10-072-621-10	Sequence 10, Appl1
19	585.5	17.6	646	US-10-072-621-9	Sequence 9, Appl1

20	585.5	17.6	646	US-10-090-455-2	Sequence 2, Appl1
21	578.5	17.4	627	US-10-090-455-8	Sequence 8, Appl1
22	578	17.4	604	US-09-745-763-197	Sequence 197, App
23	570.5	17.2	646	US-10-154-452-4	Sequence 4, Appl1
24	567.5	17.1	646	US-10-090-455-13	Sequence 13, Appl1
25	562.5	16.9	646	US-10-154-452-8	Sequence 8, Appl1
26	485	14.6	545	US-10-083-357-1335	Sequence 1335, App
27	403.5	12.1	1564	US-09-801-368-244	Sequence 244, App
28	398	12.0	1501	US-09-801-368-346	Sequence 346, App
29	374	11.2	1511	US-09-801-368-250	Sequence 250, App
30	270.5	8.1	306	US-09-769-787-83	Sequence 4554, App
31	268	8.1	242	US-09-815-242-11068	Sequence 83, Appl1
32	263.5	7.9	345	US-09-815-242-11068	Sequence 11068, A
33	263	7.3	241	US-09-815-242-11194	Sequence 11194, A
34	241.5	7.2	247	US-09-738-626-5108	Sequence 5108, App
35	240	7.2	1279	US-10-097-340-2	Sequence 2, Appl1
36	240	7.2	1280	US-10-072-621-7	Sequence 7, Appl1
37	240	7.2	1280	US-09-866-866A-2	Sequence 2, Appl1
38	240	7.2	1280	US-09-866-866A-4	Sequence 4, Appl1
39	239.5	7.2	2436	US-10-156-239-8	Sequence 8, Appl1
40	239.5	7.2	2436	US-10-199-485-8	Sequence 8, Appl1
41	239.5	7.2	2436	US-09-795-693-8	Sequence 8, Appl1
42	238.5	7.2	1594	US-09-971-121-4	Sequence 4, Appl1
43	238.5	7.2	1642	US-09-971-121-2	Sequence 2, Appl1
44	236.5	7.1	385	US-09-769-787-144	Sequence 144, App
45	236	7.1	203	US-10-252-819-17	Sequence 17, Appl1

ALIGNMENTS

RESULT 1
US-10-090-455-6
; Sequence 6, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103 406
; CURRENT APPLICATION NUMBER: US/10/090,455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-6

Query Match 100.0%: Score 3326; DB 9; Length 651;
Best Local Similarity 100.0%: Pred. No. 1.6e-285;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MODLSLTPGGSGMGLOVNRGSSQSLGAPATAPEPHSGLILHASYSVSHRVPMDITSC	60
DB	1	MGDLSTLRFGSGMGLOVNRGSSQSLGAPATAPEPHSGLILHASYSVSHRVPMDITSC	60
QY	61	ROQWROIILKDVSLVYESQIMCIIGSSGATTTLLDAMSGRAGRFTLGEVYNGRAL	120
DB	61	ROQWROIILKDVSLVYESQIMCIIGSSGATTTLLDAMSGRAGRFTLGEVYNGRAL	120
QY	121	REQOPDEFSYVLOSITLSTLYRETHLHTLALIRRCNPGSPQKYAVAAEJLSHY	180
DB	121	REQOPDEFSYVLOSITLSTLYRETHLHTLALIRRCNPGSPQKYAVAAEJLSHY	180
QY	181	ADRLIGNTSLGISTGERRRYSIAQLQDPVLMFDEPTTGLDQMTANQIVLVLELAR	240
DB	181	ADRLIGNTSLGISTGERRRYSIAQLQDPVLMFDEPTTGLDQMTANQIVLVLELAR	240
QY	241	RNRIVVLTTHQPSLRFQFDKIALISFGLIEGTPAMMDLFDNDGCPCEHSPNPDF	300
DB	241	RNRIVVLTTHQPSLRFQFDKIALISFGLIEGTPAMMDLFDNDGCPCEHSPNPDF	300

OY 301 YMDLTSVDQSKEREIEITSKRVOMIESAYKKSALCHTKNIEEMKHLKTLPMVPEKTKD 360
DB 301 YMDLTSVDQSKEREIEITSKRVOMIESAYKKSALCHTKNIEEMKHLKTLPMVPEKTKD 360
OY 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFLLFVLRVRSNVLKCAIDRV 420
DB 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFLLFVLRVRSNVLKCAIDRV 420
OY 421 GLLYQVFGATPYTGMNLAVNLPVLRVSDQESODGLYQKQOMLAVALHVLPSVYATM 480
DB 421 GLLYQVFGATPYTGMNLAVNLPVLRVSDQESODGLYQKQOMLAVALHVLPSVYATM 480
OY 481 IFSSVCYWTGLHBEVARFGFSALLAPHLIGEFLLVLLGIYQNRNIVNSVYALLSIA 540
DB 481 IFSSVCYWTGLHBEVARFGFSALLAPHLIGEFLLVLLGIYQNRNIVNSVYALLSIA 540
OY 541 GLVSGSFLNRIOEMPIPKIISYTFQKCSLLVNEFYGLNFTCGSSNVSTTTPMC 600
DB 541 GLVSGSFLNRIOEMPIPKIISYTFQKCSLLVNEFYGLNFTCGSSNVSTTTPMC 600
OY 601 AFTOGIOFIKTCPGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
DB 601 AFTOGIOFIKTCPGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 2

US-09-989-981a-6
Sequence 6, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989, 981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 651
TYPE: PR
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981a-6

Query Match 100.0%; Score 3326; DB 9; Length 651;
Best Local Similarity 100.0%; Pred. No. 1,6e-289;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGLDSLTPGSGMGLQVNRGSSSLLEGAPATAPEPHSILGILHASYSHRVRPMDTSC 60
DB 1 MGLDSLTPGSGMGLQVNRGSSSLLEGAPATAPEPHSILGILHASYSHRVRPMDTSC 60
OY 61 ROOMTRQILKDVSLVYESGOIMCILLSSGSGKTTLLDAMSGRLGRAGTFIGEYVNGRAL 120
DB 61 ROOMTRQILKDVSLVYESGOIMCILLSSGSGKTTLLDAMSGRLGRAGTFIGEYVNGRAL 120
OY 121 RRQFQDCFSYVLOSOTLLSLVRETHTALLAIRGNPGSFQKVEVMAELSLSHV 180
DB 121 RRQFQDCFSYVLOSOTLLSLVRETHTALLAIRGNPGSFQKVEVMAELSLSHV 180
OY 181 ADRIGNYSLGISTGERRRVSIQAOLDPKVMLEDEPTGLDCMTANOIVLVLELAR 240
DB 181 ADRIGNYSLGISTGERRRVSIQAOLDPKVMLEDEPTGLDCMTANOIVLVLELAR 240

DB 181 ADRIGNYSLGISTGERRRVSIQAOLDPKVMLEDEPTGLDCMTANOIVLVLELAR 240
OY 241 RNRIVLTLTHQPRSELEFQEDKIALISFGEILFCGTPAEMLDEFENDCGYPCPEHSNPDF 300
DB 241 RNRIVLTLTHQPRSELEFQEDKIALISFGEILFCGTPAEMLDEFENDCGYPCPEHSNPDF 300
OY 301 YMDLTSVDQSKEREIEITSKRVOMIESAYKKSALCHTKNIEEMKHLKTLPMVPEKTKD 360
DB 301 YMDLTSVDQSKEREIEITSKRVOMIESAYKKSALCHTKNIEEMKHLKTLPMVPEKTKD 360
OY 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFLLFVLRVRSNVLKCAIDRV 420
DB 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFLLFVLRVRSNVLKCAIDRV 420
OY 421 GLLYQVFGATPYTGMNLAVNLPVLRVSDQESODGLYQKQOMLAVALHVLPSVYATM 480
DB 421 GLLYQVFGATPYTGMNLAVNLPVLRVSDQESODGLYQKQOMLAVALHVLPSVYATM 480
OY 481 IFSSVCYWTGLHBEVARFGFSALLAPHLIGEFLLVLLGIYQNRNIVNSVYALLSIA 540
DB 481 IFSSVCYWTGLHBEVARFGFSALLAPHLIGEFLLVLLGIYQNRNIVNSVYALLSIA 540
OY 541 GLVSGSFLNRIOEMPIPKIISYTFQKCSLLVNEFYGLNFTCGSSNVSTTTPMC 600
DB 541 GLVSGSFLNRIOEMPIPKIISYTFQKCSLLVNEFYGLNFTCGSSNVSTTTPMC 600
OY 601 AFTOGIOFIKTCPGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
DB 601 AFTOGIOFIKTCPGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 3

US-09-837-992-3
Sequence 3, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bel
APPLICANT: Tularik Inc.
APPLICANT: Sltosterolemia Susceptibility Gene (SSG): Compositions
TITLE OF INVENTION: and Methods of Use
FILE REFERENCE: 018781-006020US
CURRENT APPLICATION NUMBER: US/09/837, 992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 651
TYPE: PR
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
US-09-837-992-3

Query Match 100.0%; Score 3326; DB 10; Length 651;
Best Local Similarity 100.0%; Pred. No. 1,6e-289;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGLDSLTPGSGMGLQVNRGSSSLLEGAPATAPEPHSILGILHASYSHRVRPMDTSC 60
DB 1 MGLDSLTPGSGMGLQVNRGSSSLLEGAPATAPEPHSILGILHASYSHRVRPMDTSC 60
OY 61 ROOMTRQILKDVSLVYESGOIMCILLSSGSGKTTLLDAMSGRLGRAGTFIGEYVNGRAL 120
DB 61 ROOMTRQILKDVSLVYESGOIMCILLSSGSGKTTLLDAMSGRLGRAGTFIGEYVNGRAL 120
OY 121 RRQFQDCFSYVLOSOTLLSLVRETHTALLAIRGNPGSFQKVEVMAELSLSHV 180
DB 121 RRQFQDCFSYVLOSOTLLSLVRETHTALLAIRGNPGSFQKVEVMAELSLSHV 180

121 RRRQFDDCSYVLOSOTLLSLVRETLHTALLARKNGPSFOKKEAVALSLSH 180
DB
181 ADRLLGNYSLAGISTGERRRVSIQAOLLODPKVMLEDEPTTGDCMTANOIVLVEIAR 240
QY
181 ADRLLGNYSLAGISTGERRRVSIQAOLLODPKVMLEDEPTTGDCMTANOIVLVEIAR 240
DB
241 RRRIVVLTTHQPRSELFOLFDKIALISFGEILFCGTPAEMLDFNDCGYPCEHSNPDF 300
QY
241 RRRIVVLTTHQPRSELFOLFDKIALISFGEILFCGTPAEMLDFNDCGYPCEHSNPDF 300
DB
301 YMDLTSVDTOSKEREIETSKRVOMIESAVKSAICHTLKNIERMKHTLTPMVPKTKD 360
QY
301 YMDLTSVDTOSKEREIETSKRVOMIESAVKSAICHTLKNIERMKHTLTPMVPKTKD 360
DB
361 SPQVSKLGLVLLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVRSNVLKGAIQDRV 420
QY
361 SPQVSKLGLVLLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVRSNVLKGAIQDRV 420
DB
421 GLLYQVGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKQMMLAYALHVPESVATM 480
QY
421 GLLYQVGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKQMMLAYALHVPESVATM 480
DB
421 GLLYQVGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKQMMLAYALHVPESVATM 480
QY
481 IFSSVCYMTLGLHPEVARGYFSALLAHLIGEFLTLVLGIYONPNIVNSVVALLSIA 540
DB
481 IFSSVCYMTLGLHPEVARGYFSALLAHLIGEFLTLVLGIYONPNIVNSVVALLSIA 540
QY
541 GVLVSGFLRNIOEMPIPKIISYTFQYKCEIIVNVEFYGLNFCGSSNVSVTNPMK 600
DB
541 GVLVSGFLRNIOEMPIPKIISYTFQYKCEIIVNVEFYGLNFCGSSNVSVTNPMK 600
QY
601 AFTOGIOFIEKTCPGATSRFTMNFLLYSFIPALVILGIVFKIRDLISR 651
DB
601 AFTOGIOFIEKTCPGATSRFTMNFLLYSFIPALVILGIVFKIRDLISR 651

Result 4

US-09-989-981A-2

Sequence 2, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse ABCG5 (MABG5)
US-09-989-981A-2

Query Match 82.5%; Score 2744.5; DB 9; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
QY 1 MGDLSSLPFGSGMGLQVNGSOSLEGAPATAPED-HSIGILHAHAYSHRVPRMMDITS 59
DB 1 MGELFLPSPEGARGHINSGSLSLQSGSVTCTEARRHSLGVLHAYSVSNRKGPMWNIS 60

60 CRQOATROLKDVSLVYESGOIWCITLSSGSGKTTLLDAMSGFLGAGFFLGEVYNGR 119
DB
61 COOKMDROILKDVSLVIESGOIWCITLSSGSGKTTLLDAMSGFLGAGFFLGEVYNGR 120
QY
120 LRREPODCESVYLOSOTLLSLVRETLHTALLARKNGPSFOKKEAVALSLSH 179
DB
121 LRREPODCESVYLOSOTLLSLVRETLHTALLARKNGPSFOKKEAVALSLSH 180
QY
180 VADRLLGNYSLAGISTGERRRVSIQAOLLODPKVMLEDEPTTGDCMTANOIVLVEIAR 239
DB
181 VADRLLGNYSLAGISTGERRRVSIQAOLLODPKVMLEDEPTTGDCMTANOIVLVEIAR 240
QY
240 RRRIVVLTTHQPRSELFOLFDKIALISFGEILFCGTPAEMLDFNDCGYPCEHSNPDF 299
DB
241 RRRIVVLTTHQPRSELFOLFDKIALISFGEILFCGTPAEMLDFNDCGYPCEHSNPDF 300
QY
300 FYMDLTSVDTOSKEREIETSKRVOMIESAVKSAICHTLKNIERMKHTLTPMVPKTKD 359
DB
301 FYMDLTSVDTOSKEREIETSKRVOMIESAVKSAICHTLKNIERMKHTLTPMVPKTKD 360
QY
360 DSEGVFSKGLVLLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVRSNVLKGAIQDRV 419
DB
361 DSEGVFSKGLVLLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVRSNVLKGAIQDRV 420
QY
420 VGLLYQVGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKQMMLAYALHVPESVAT 479
DB
421 VGLLYQVGAFTPTGMLNANVLEPVLRAVSDQSDGLYOKQMMLAYALHVPESVAT 480
QY
480 MIFSSVCYMTLGLHPEVARGYFSALLAHLIGEFLTLVLGIYONPNIVNSVVALLSI 539
DB
481 MIFSSVCYMTLGLHPEVARGYFSALLAHLIGEFLTLVLGIYONPNIVNSVVALLSI 540
QY
540 AGVLVSGFLRNIOEMPIPKIISYTFQYKCEIIVNVEFYGLNFCGSSNVSVTNPMK 599
DB
541 SGLLISGFLRNIOEMPIPKIISYTFQYKCEIIVNVEFYGLNFCGSSNVSVTNPMK 600
QY
600 CAFTOGIOFIEKTCPGATSRFTMNFLLYSFIPALVILGIVFKIRDLISR 651
DB
601 CAFTOGIOFIEKTCPGATSRFTMNFLLYSFIPALVILGIVFKIRDLISR 652

Result 5

US-09-837-992-1

Sequence 1, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Stossterolemia Susceptibility Gene (SSG): Compositions
FILE REFERENCE: 018781-006020US
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse stossterolemia susceptibility gene (SSG)
US-09-837-992-1

Query Match 82.5%; Score 2744.5; DB 10; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

Query Match	21.0%;	Score 697;	DB 9;	Length 673;
Best Local Similarity	28.9%;	Pred. No. 9.5e-54;		
Matches 187;	Conservative 124;	Mismatches 241;	Indels 96;	Gaps 16;

```

0Y 8 TPGSGAGLOVNGNSOSSLGCAVPT -APEHSIGLTHASVSHRV -PMWD -ITSQOOW 64
   ||: ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
16 TPQDTSGLODRLEFSSSDNSLYFTTYSGQPTLEVRLOINTVODIASOVPMWEOLAQOFMPW 75
0Y 65 TROI-----LKQDVLYESGQIMCIGSSGSGTTLTDAMSGRLGRGT- LGEVY 115
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 TSPSCONSCLEIQMISFKYVRSGQMALIGSSCGRASILDVITGR -GHGKIKSGIWI 134
0Y 116 NSRALRREGFOCFEVSUOJSDLLSLSTVRETIHYTALLAI -RKGNGSQRKYEAVMAE 174
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 135 NQOPSSPOLVRCVAVHROHNDLPLVRETIATLAQMKLPTFFSOAQDRKREVDIAE 194
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
0Y 175 LSLSHVADRLIGNYSIGIGISTGERRRVSTIAAOLLOPKWALFDEPTTGDCMTANOIVL 234
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 195 LRLRQCADTRVGNMVMYRGISGGERRRVSTIGVOLNNPGLIIDEPTSGDLSFAHNLVKT 254
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
0Y 235 LVELARRNRIVLTHQPSSELFOLEFDKTAIISFGLIFCGPRAEMIDFENDGYPDEH 294
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 255 LSRLAKGNLVYISLHOPRSDIFRLEPDLVLLTSGPIYLGAAQHWQYFTAJPCPRY 314
0Y 295 SNPFDFYMDLTVSDNOSKREIETSKRYOMITSAYKKSALCHTLANIEMKML----- 348
   ||| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : :
Db 315 SNRPAFYDVTISIDRRSKRQELATREKAOISLALF-----LEEVRLODFLMK 362
0Y 349 -----FTLPM-----VPEKTSBGEVFSKLGVLRLRYTRNLVNRMLAVTTL 390
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 363 AETKDLDEDTCEVSESVTPLDTCNPCLSPTR -MGAVQOFTTLIRQISNDRDLPTLLIHG 421
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
0Y 391 LQNLIMGLELFEVLARVSNVLKGAIQ-----DRVGLIYQVCGATPYTGMALNVLFPVLR 446
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 422 AEACIMSMTIGLRYG-----HGSIOJSEMDTALLFMGALIPRNVILDIIVSCYSESR 475
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
0Y 447 AVSDQESDGLQOKQOMLALYALHVLPEFSVATMTFSSVCYMTLGLHPEVAR----- 499
   ||: ||: ||| ||| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 476 AMLYELBDGLTTPPYFPAKLTIGELPEICAIITIIIGMPTLMIANLRPGLOPFLHFLV 535
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
0Y 500 -----GTFSAALLAPHLIGELTVLLGIQVONPVIIVNSVALLSIAGVLVSGGFL 549
   ||: ||| ||| : : : : : : : : : : : : : : : : : : : : : :
Db 536 WLNVFCCRIMALAALLPTFHMASEFS-----NALYNSFYLAG-----GFM 577
0Y 550 RNIQEMPIPKLISYTFQKOCSELIIVNAEFGLMPTCGSSNVSVYTIN 597
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 578 INLSMLTVPAIVSIVSFLRWCFEGLMKIQESRRRYTKMDLGNITIAVS 625

```

RESULT 8
US-09-989-981A-8
: Sequence 8, Application US/09989981A
: Publication No. US20030049730A1
: GENERAL INFORMATION:
: APPLICANT: Hobbs, Helen H.
: APPLICANT: Shan, Bel
: APPLICANT: Barnes, Robert
: APPLICANT: Tian, Hui
: APPLICANT: Tularik Inc.
: TITLE OF INVENTION: Board of Regents, The University of Texas System
: FILE REFERENCE: 018781-00732005
: CURRENT APPLICATION NUMBER: US/09/989,981A
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: US 60/252,235
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: US 60/253,645
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 673

```

: TYPE: prt
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: OTHER INFORMATION: human ABCG8 (hABCG8)
:
US-09-989-981A-8

```

Query Match	21.0%;	Score 697;	DB 9;	Length 673;
Best Local Similarity	28.9%;	Pred. No. 9.5e-54;		
Matches 187;	Conservative 124;	Mismatches 241;	Indels 96;	Gaps 16;

QY	8	TPGSGMGLVNNKSGSSLEGAPAT-APPHSAGILHMSYSHRYR-PMMD-ITSCROOW	64
Dd	16	TPQDTSGLDRLFFSSSDNSLXFTYSQGPNTLEVRDNTQYDASLQYPMPEQLAQFKNPM	75
QY	65	TRQI-----LKDVSLYESGQIMCILLGSSGSKTTLDAMSGRUGRGTF-IGEVY	115
QY	116	NGPALARPEODCFSVLSQSDPILLSLVRETNHTALL-REGNGCSQOKKYEAMAE	174
Dd	76	TPSPCONSCELQIOMLSEFKVRSGOMALIIIGSSCGRASLADIVIGR-GHGKIKSGIWI	134
Dd	135	NGQPSSPOLYRKCVAHVROHNOPLLIVRETLFAIQMRIPRFSSQARKRVEDVIAE	194
QY	175	LSLSHVADRLIGNYSLGISTGERRRYSIAAOLDDPKMLDEPTTGLCMTANQIYVL	234
Dd	195	LRLRQCADTRVGNMVRGSGGERRRVSIGVOLLMPGILLDDEPTGSDSFTAHNLVKT	254
QY	235	LYELARRNIYVLTTHQRPSELFOJDKIALISFQELIFCGTPEMLDFENDCGYPCPEH	294
Dd	255	LSRLAKGNLVLISLHOPRSDIFRLFDLYVLLTSGTPYLGAAOHMVOYFALIGPCPRY	314
QY	295	SNPFDEYMLDTSVDOSKERETETSKRYOMIESAYKKSALCHTKLNIEPMKH-	348
Dd	315	SNRPADPYDUTSIDRKSREQELATEKQSLAALF-----LEVRDDLDQFLMK	362
QY	349	-----KTLPM-----VPEKTKDSPGVFSKGLVLRREVTENLVNKLAVTRL	390
Dd	363	AETKDLDDEDTCESSVTPLEDTNCLPSPK-MPGAVQOFTTLIRQISNDPRDLPTLLIHG	421
QY	391	LQNLIMGFLFLEFVLVRASNVLKALQ-----DRGLLYQYVGATPYTGMLNAAVLPVLR	446
Dd	422	AELCAISMSTIGLYFG-----HSGIQLSFMDTALLFMIGALLPFNVILLDVISKCYSER	475
QY	447	AVSDSEODQGLQOKOMMLAYALAHVLPESVAYATIFSSVCWTTGLLPEVARF-----	499
Dd	476	AMLYYELDEGLYTTQPYFPKILGELPBHCAYIIITIGMPYTWLANLRPGLOPELLHLLY	535
QY	500	-----GYESALLAPHLIGBELTLLVLLGIVQNPNTVNSVALLSIAGYLVYSGFL	549
Dd	536	WLVEVCCRIMALAAALLPTFHMASSF-----NALYNSFYLAG-----GFM	577
QY	550	RNIQEMPIPKIISYTFQKCYSELVYNEGYLGNFCGSSNNSVYTN	597
Dd	578	ILUSSITVWPAMISVSEFLKMCFGKLMKIQSSRRTPYKPLGNTLIAVS	625

RESULT 9
 US-09-961-086-1
 ; Sequence 1, Application US/09961086
 ; Publication No. US20030036645A1
 ;
 GENERAL INFORMATION:
 ; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
 ; APPLICANT: ROSS, Douglas D.
 ; APPLICANT: DOYLE, L. Austin
 ; APPLICANT: ABRUZZO, Lynne
 ; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
 ; TITLE OF INVENTION: WHICH ENCODES IT
 ; FILE REFERENCE: EP19376-019
 ; CURRENT APPLICATION NUMBER: US/09/961,086
 ; CURRENT FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 60/073,763
 ; PRIOR FILING DATE: 1998-02-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/02577
 ; PRIOR FILING DATE: 1999-02-05

NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-09-981-086-1

Query Match 20.5%; Score 682.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 1.8e-52;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

```
OY 21 SSSLEGAPATAP---EPHSLGILHASVSASHRRPMMMDITSCQOQTRQLKQVSLVYE 77
   ||: || || || | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 SGGNTGFPATASNDLKAFTGCAVLSFNICYRVKLSGGLPCPKPVEKELISNIGIMK 72
OY 78 SGOIMCIGSSGSGKTTLLDAMSGRLAGCTFLEEVYNGRALLREOFQDFSVYLOSDT 137
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 73 PG-LNALIGPTGGKSSLDVLAARKDPSG-LSGDVILNG-APPAPAFKCSGYVDDV 129
OY 138 LLSLTYRETLHYTALLAIRGNPG-SFOKKEVAEMELSLSHVADRLIGNYSLGISTG 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 VMGLTYRENLQFSAALRLATTTNHEKNERINRYIOELGIDKVAADSKVGTQIRVSGG 189
OY 197 ERRRVSTAAOLLQDPKVMLEDEPTTGIDCMQANQIYVLVLELARRNRYVLTIQPSEL 256
   ||: || | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 190 ERKRTSIGMELITDPSILFDEPTGLDSSSTANAVLLLLKRMKOGRTIIFSHQPRYSI 249
OY 257 FQLEFDKIALISFGELIFCGFPAEMLDFPNDGYPCEPHSNPFDYMDLTSVDTO-----SK 312
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 250 FKLDSLTLLASGRMLHGPQOELAGYFESAGYCEAVNNADPFDLIINDSTAVALLNR 309
OY 313 ERE-----IETSR---VOMIESAYKSAICKT-----LKNIERMKHLKTLPMVPF 356
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 310 EEDKATEIIEPSSKODRPLEKLAIEIYVNSSEFYKETAELHQLSGEKKKITYEKEISY 369
OY 357 KTKDSPGVESKGLVLRVRNRYRNKLAIVTLLQNLINGLFL--LFFVLRVSNVLKG 414
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 370 TT-----SFCQLRNVSKSRKFNKLNQNPQASTAQIITVYVGLVIGAIYFGKLNST---- 421
OY 415 AIODRVGLLYQFVATPYTGMNAVNLEPVLRAVSDDESODGLYOKQOMLAVAL-HVLP 473
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 422 GIORAGVLF-FLTNGCFSSVSVAVELFYVEKKLFHEIYISGYRVSSYFLGKLSLDLP 480
OY 474 FSVYATMIFSSVCWTGLGHPVARFGYFSNALLAPHLIGELTLLVLLGIVONPNIVSV 533
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 481 MRMPLSIITFCIVFEMGLKPKADAFVMMFTLM---MVAYSASSMALAIAGAOSVVA 537
OY 534 VALISAGV--LVSSGLRNIOEMPIPKIISYTFPOKYCEILVYNEFYGLNFTCCSSN 591
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 538 TLMATICFPMIIFSGLLVNLTTASWLSWLYFSIPRYGFTALQHNHFLGQNECPG--- 594
OY 592 VSVTTNPMCAFTOGIOFIEKTCPG 615
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 595 LMATGNMPCNYA-----TCTG 610
```

RESULT 10
US-09-981-353-35
Sequence 35, Application US/09981353
Patent No. US20020160382A1

GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
APPLICANT: Jones, David A.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981.353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 35
LENGTH: 655
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

Query Match 20.5%; Score 680.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.7e-52;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

```
OY 21 SSSLEGAPATAP---EPHSLGILHASVSASHRRPMMMDITSCQOQTRQLKQVSLVYE 77
   ||: || || || | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 SGGNTGFPATASNDLKAFTGCAVLSFNICYRVKLSGGLPCPKPVEKELISNIGIMK 72
OY 78 SGOIMCIGSSGSGKTTLLDAMSGRLAGCTFLEEVYNGRALLREOFQDFSVYLOSDT 137
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 73 PG-LNALIGPTGGKSSLDVLAARKDPSG-LSGDVILNG-APPAPAFKCSGYVDDV 129
OY 138 LLSLTYRETLHYTALLAIRGNPG-SFOKKEVAEMELSLSHVADRLIGNYSLGISTG 196
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 130 VMGLTYRENLQFSAALRLATTTNHEKNERINRYIOELGIDKVAADSKVGTQIRVSGG 189
OY 197 ERRRVSTAAOLLQDPKVMLEDEPTTGIDCMQANQIYVLVLELARRNRYVLTIQPSEL 256
   ||: || | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 190 ERKRTSIGMELITDPSILFDEPTGLDSSSTANAVLLLLKRMKOGRTIIFSHQPRYSI 249
OY 257 FQLEFDKIALISFGELIFCGFPAEMLDFPNDGYPCEPHSNPFDYMDLTSVDTO-----SK 312
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 250 FKLDSLTLLASGRMLHGPQOELAGYFESAGYCEAVNNADPFDLIINDSTAVALLNR 309
OY 313 ERE-----IETSR---VOMIESAYKSAICKT-----LKNIERMKHLKTLPMVPF 356
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 310 EEDKATEIIEPSSKODRPLEKLAIEIYVNSSEFYKETAELHQLSGEKKKITYEKEISY 369
OY 357 KTKDSPGVESKGLVLRVRNRYRNKLAIVTLLQNLINGLFL--LFFVLRVSNVLKG 414
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 370 TT-----SFCQLRNVSKSRKFNKLNQNPQASTAQIITVYVGLVIGAIYFGKLNST---- 421
OY 415 AIODRVGLLYQFVATPYTGMNAVNLEPVLRAVSDDESODGLYOKQOMLAVAL-HVLP 473
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 422 GIORAGVLF-FLTNGCFSSVSVAVELFYVEKKLFHEIYISGYRVSSYFLGKLSLDLP 480
OY 474 FSVYATMIFSSVCWTGLGHPVARFGYFSNALLAPHLIGELTLLVLLGIVONPNIVSV 533
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 481 MRMPLSIITFCIVFEMGLKPKADAFVMMFTLM---MVAYSASSMALAIAGAOSVVA 537
OY 534 VALISAGV--LVSSGLRNIOEMPIPKIISYTFPOKYCEILVYNEFYGLNFTCCSSN 591
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 538 TLMATICFPMIIFSGLLVNLTTASWLSWLYFSIPRYGFTALQHNHFLGQNECPG--- 594
OY 592 VSVTTNPMCAFTOGIOFIEKTCPG 615
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 595 LMATGNMPCNYA-----TCTG 610
```

RESULT 11
US-10-120-687-61
Sequence 61, Application US/10120687
Publication No. US20030082155A1

GENERAL INFORMATION:
APPLICANT: Massachusetts General Hospital
TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treat
FILE REFERENCE: 3284/1235B
CURRENT APPLICATION NUMBER: US/10/120.687
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: US60/169082
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: US 09/963,875
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/215109
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: US 60/238880

; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: US 09/731261
 ; PRIOR FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 61
 ; LENGTH: 655
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-120-687-61

Query Match 20.5%; Score 680.5; DB 9; Length 655;
 Best Local Similarity 29.2%; Pred. No. 2,7e-52;
 Matches 182; Conservative 137; Mismatches 220; Indels 55; Gaps 18;

QY 21 SOSLEGAPATAP---EPHSIGILHASYSVSHRVRPMMWDTSCROQMTROLKXVLYE 77
 DB 13 SGGTNGFPATVSDKLAFTGAVLSFHNTICRYKLKSGFLPCRKPEKEILSNINGIMK 72
 QY 78 SGOIMCIISSGSGKTTLLDAMSGRLGACFTFLGEEVYVNGRALRREOFODCFSTVYLOSDT 137
 DB 73 PG-LNALIGPTGGKSSLDVLARKDPSG-LSGDVILNG-APRPAFKCKSGYVODDV 129
 QY 138 LLSLTYRETLHATALLAIRGNPG-SFOKKVEAVMAELSLSHVADRLIGNYSLGISTG 196
 DB 130 VMGLTYRENLQFSAALRLATYTNHNEKNERINRYIELGLDKVADSKVGTQFTRGYSGG 189
 QY 197 ERRRVSTIAOOLLQDPKWLDEPTTGLDCMTANOIVLVLELARNRNIYVLTTHQPSSEL 256
 DB 190 EKRRTSIGMELITDPSILFDEPTTGLDSTANAVLILLKRMKSGRTIIFSHQPRSI 249
 QY 257 FOLFDKTALLSFGELICGTPAEMLDPEFNDGYPCEPHSNPFDEYMDLTSVDTO----SK 312
 DB 250 FKLPDLSLTLLASGRMLFHGPAQALGYFESAGYCEAYNNPADFELDIINGDSTAVALNR 309
 QY 313 ERE-----IETSKR-----VOMIESAVKSAICHKT-----LKNIERMKHLKTLPAVPF 356
 DB 310 EEDPKATEIIEPSKODPLIEKLAETIVNSFYETKAEHLQSLGGEKKKIYFKEISY 369
 QY 357 KTKDSPGVESKLGVLRRVTRNRLAVITRLLQMLNGLEL--LEFVLVRANSYVLK 414
 DB 370 TT-----SFCHQLRNVSRKSRKNNLGNQASIAQIIVTVGLVIGALYFGKNDST---- 421
 QY 415 AIDRVGLLYQFVATPYTGMLNANVLPVLAASVDSODGLYOKQMMALAYAL-HVLP 473
 DB 422 GIONRAGVLP-FLITNOCFSSVSVELFVEYKELFHEIYISGYRVSSEYFLGKLSDLLP 480
 QY 474 FSVVATMIFSSVCYTWLGLHPEVARGFYFSALLAPHLIGBFLTVLLGIVQNPNIYNSV 533
 DB 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTLL--MVAYSASSMALALAIAGOSVVA 537
 QY 534 VALISTAGV--LVGSGFLRNIOEMPIPKIISYFTPOKYSSEILVNEFGLNFTCCSSN 591
 DB 538 TLMTATCFVEMMIFSGLLVNLTTIASWLSMLOYFSIRYGFYALQHNHPEFGONCPG--- 594
 QY 592 VSVTTNPMCAFTGIGIETKCPG 615
 DB 595 LNAATGNMPCMYA-----TCYG 610

RESULT 12
 US-10-090-455-5
 ; Sequence 5, Application US/10090455
 ; Publication No. US20030027259A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Hongyun
 ; APPLICANT: Le Bihan, Stephane
 ; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
 ; FILE REFERENCE: 100103.406
 ; CURRENT APPLICATION NUMBER: US/10/090.455
 ; PRIOR FILING DATE: 2002-03-01
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5
 ; LENGTH: 655
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-090-455-5

Query Match 20.3%; Score 674.5; DB 9; Length 655;
 Best Local Similarity 29.0%; Pred. No. 9.5e-52;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SOSLEGAPATAP---EPHSIGILHASYSVSHRVRPMMWDTSCROQMTROLKXVLYE 77
 DB 13 SGGTNGFPATVSDKLAFTGAVLSFHNTICRYKLKSGFLPCRKPEKEILSNINGIMK 72
 QY 78 SGOIMCIISSGSGKTTLLDAMSGRLGACFTFLGEEVYVNGRALRREOFODCFSTVYLOSDT 137
 DB 73 PG-LNALIGPTGGKSSLDVLARKDPSG-LSGDVILNG-APRPAFKCKSGYVODDV 129
 QY 138 LLSLTYRETLHATALLAIRGNPG-SFOKKVEAVMAELSLSHVADRLIGNYSLGISTG 196
 DB 130 VMGLTYRENLQFSAALRLATYTNHNEKNERINRYIELGLDKVADSKVGTQFTRGYSGG 189
 QY 197 ERRRVSTIAOOLLQDPKWLDEPTTGLDCMTANOIVLVLELARNRNIYVLTTHQPSSEL 256
 DB 190 EKRRTSIGMELITDPSILSDEPTTGLDSTANAVLILLKRMKSGRTIIFSHQPRSI 249
 QY 257 FOLFDKTALLSFGELICGTPAEMLDPEFNDGYPCEPHSNPFDEYMDLTSVDTO----SK 312
 DB 250 FKLPDLSLTLLASGRMLFHGPAQALGYFESAGYCEAYNNPADFELDIINGDSTAVALNR 309
 QY 313 ERE-----IETSKR-----VOMIESAVKSAICHKT-----LKNIERMKHLKTLPAVPF 356
 DB 310 EEDPKATEIIEPSKODPLIEKLAETIVNSFYETKAEHLQSLGGEKKKIYFKEISY 369
 QY 357 KTKDSPGVESKLGVLRRVTRNRLAVITRLLQMLNGLEL--LEFVLVRANSYVLK 414
 DB 370 TT-----SFCHQLRNVSRKSRKNNLGNQASIAQIIVTVGLVIGALYFGKNDST---- 421
 QY 415 AIDRVGLLYQFVATPYTGMLNANVLPVLAASVDSODGLYOKQMMALAYAL-HVLP 473
 DB 422 GIONRAGVLP-FLITNOCFSSVSVELFVEYKELFHEIYISGYRVSSEYFLGKLSDLLP 480
 QY 474 FSVVATMIFSSVCYTWLGLHPEVARGFYFSALLAPHLIGBFLTVLLGIVQNPNIYNSV 533
 DB 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTLL--MVAYSASSMALALAIAGOSVVA 537
 QY 534 VALISTAGV--LVGSGFLRNIOEMPIPKIISYFTPOKYSSEILVNEFGLNFTCCSSN 591
 DB 538 TLMTATCFVEMMIFSGLLVNLTTIASWLSMLOYFSIRYGFYALQHNHPEFGONCPG--- 594
 QY 592 VSVTTNPMCAFTGIGIETKCPG 615
 DB 595 LNAATGNMPCMYA-----TCYG 610

RESULT 13
 US-09-866-10
 ; Sequence 10, Application US/09866866A
 ; Patent No. US20020102244A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sorrentino, Brian
 ; APPLICANT: Schuetz, John
 ; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
 ; FILE REFERENCE: 1340-1-021CIP2
 ; CURRENT APPLICATION NUMBER: US/09/866.866A
 ; PRIOR FILING DATE: 2001-08-30
 ; PRIOR APPLICATION NUMBER: 09/584,586
 ; PRIOR FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: PCT/US99/11825
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: 60/086,988
 ; PRIOR FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 27

PRIOR APPLICATION NUMBER: 60/086,988
PRIOR FILING DATE: 1998-05-28

PRIOR APPLICATION NUMBER: PCT/US99/11825
PRIOR FILING DATE: 1999-05-27

PRIOR APPLICATION NUMBER: PCT/US99/11825
PRIOR FILING DATE: 1999-05-27

;; PRIOR FILING DATE: 1998-05-28
;; NUMBER OF SEQ ID NOS: 27
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 14
;; LENGTH: 657
;; TYPE: PRT
;; ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 19.88; Score 660; DB 10; Length 657;
Best Local Similarity 28.08; Pred. No. 1.9e-50;
Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

```

QY 13 MGLQVNRGSSQSLGAPATAPAPAPHSIGLILASTSVSHRVPRPMWDITSCROQWTROILKDY 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 MSQRNNNGLPBRMSRAVRITLAEGLVSEFHITTRV--KXSGFLV---RKTVEKEILLSDI 66
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 SLVESGQIMCIIIGSSGSGKTTLLDAMSGRLGRAGTELEEVYNGRALRREOPQDCFSYV 132
   : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 NGIMKPG-LNAILGPTGGCKSLDLVLAARKDKPG-LSGDVLNG--APQAHFKCCSGYV 123
   : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 LQSDTLLSLTVEETLHYTALLAIRGNQSGFO-----KKEAVMAELSLSHVADRILGN 187
   : | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 VQDDVVGTLTVRENLOFSAALRL---PTMKNHKERNERTIIEKELGEEKVADSKVGP 179
   : | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 YSLGISTGERRRVSIAGLADPKYMLFEPTTGLDCMTANOIVLIVLVELARRNRIVVL 247
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 QFIRGISGGRKRTSIGMELITPSTILFDEPTTGLDSSSTANAVLLLRKMSKQGRITIF 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 THQPRSELEFQLEFKIALISFGEELIFCGTAPAEMLDEFNDGYPCEPHSNPFDFMDLTSV 307
   : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 240 SIHQPRYSIFKLEPDSLTLASGLVPHGPAQKALEYFASAGYCEPYNNPADFEFLDYING 299
   : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
QY 308 DTQS-----KERIETSKR-----VQMTESAIVKKSATICHKTLKNIERNKHILKLP 352
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 300 DSSAVMLNREEDNEANKTEEPSKGEKPYENLSEFYINSATYGETKAEIDLQ----- 352
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
QY 353 MVEFKTKDSPGVFSKLGV-----LRRVPRNLVRNKLAVITRLLQNL 394
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 353 -----PGAQEKKGTSAFKEPEYVYTSFCHQLRWIARRSEFKNLGNPOASVAQOLITV 403
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 395 IMGLFL--LEFVLRRVSNVLKGAIDRVGLLYQVGFATPYTGMLNANVLEPVLRAVSDOE 452
   | : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : |
Db 404 ILGLIIGAIYFDLKYA---AGMQRACVLF-FLTTNOCFSSVSAVELEFYVEKKLFIEH 458
   | : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : |
QY 453 SDDGLYQKQMMALATL-HVLPFSVATMIFSSVCYWTGLGHEVARFGYFSALLAPHL 511
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 YISGYRVSSYFEGKVMSDLPLPRFPLPSVIFTCILFEMGLKKTVDADFIMMFTLI--M 515
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 512 IGEFLTLVLLGIYQNFNIYNSVALISTGV--LVGSGFLRNIOEMPIFKIISYTFQK 569
   | : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : |
Db 516 VAYTASSMALAIATGQSVSVATLMTTIAFVEMMLPFSGLLVNLRTIGPMLSMLOYFSIDR 575
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 570 YCSEILVNEFYGLNFTCCSSNVSVTTPMCAFTOGIOFIKTCPCG 615
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 576 YGFTALQYNEFLGQEFPCG---FNVTDNCTCVNSYAI-----CTG 612
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: June 11, 2003, 09:14:46
Job time : 26 secs

XX WPI: 2002-416483/44.
 DR N-PSDB: ABR51681.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT strossterolemia, arteriosclerosis and heart diseases
 PS Claim 52; Page 35-36; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing strossterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX
 SQ Sequence 651 AA:
 Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGDISSITPGSGMGLQVNRSGSSLEGAPATAPAPPHSLGILHASYSVSHRVRPMDITSC 60
 DB 1 MGDISSITPGSGMGLQVNRSGSSLEGAPATAPAPPHSLGILHASYSVSHRVRPMDITSC 60
 QY 61 RQOKTROIILNDVSLYVSSGOIMCTIGSSGSKTLLDAMSGRGRAGTFFGEYVNRAL 120
 DB 61 RQOKTROIILNDVSLYVSSGOIMCTIGSSGSKTLLDAMSGRGRAGTFFGEYVNRAL 120
 QY 121 RREQFOQCFYVLOSPTLLSLIYRETLHTALAIRGNPGSFQKVEYVMAELSLSHV 180
 DB 121 RREQFOQCFYVLOSPTLLSLIYRETLHTALAIRGNPGSFQKVEYVMAELSLSHV 180
 QY 121 RREQFOQCFYVLOSPTLLSLIYRETLHTALAIRGNPGSFQKVEYVMAELSLSHV 180
 DB 121 RREQFOQCFYVLOSPTLLSLIYRETLHTALAIRGNPGSFQKVEYVMAELSLSHV 180
 QY 181 ADRLIGNVSLAGISGERRRVSINAQILLDQPKYVLFDEPTTGLDCMTANOIVVLVELAR 240
 DB 181 ADRLIGNVSLAGISGERRRVSINAQILLDQPKYVLFDEPTTGLDCMTANOIVVLVELAR 240
 QY 241 RNRIVVLTIHOPRSELFOLFDKTAIISFGELIFCGTAAEMLDFFNDCGCPRESNPFDF 300
 DB 241 RNRIVVLTIHOPRSELFOLFDKTAIISFGELIFCGTAAEMLDFFNDCGCPRESNPFDF 300
 QY 301 YMDLTVDYOSKEREIESTRKVMIESAYKKSACHTLKNIEREMKILKTLPMVPEKTD 360
 DB 301 YMDLTVDYOSKEREIESTRKVMIESAYKKSACHTLKNIEREMKILKTLPMVPEKTD 360
 QY 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFFLFLVLRVSNVLGAIDRV 420
 DB 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTRLLQNLIMGLFFLFLVLRVSNVLGAIDRV 420
 QY 421 GLIYOFGATPYTGMLAANVLFVTLRAVSDQSDGLYQKWMMLAALAVLDFSVATM 480
 DB 421 GLIYOFGATPYTGMLAANVLFVTLRAVSDQSDGLYQKWMMLAALAVLDFSVATM 480
 QY 481 IFSSVCYWTGLGHEVARFGYFSAALAPHLIGFELTVLGIYQNPINYSVVALLSIA 540
 DB 481 IFSSVCYWTGLGHEVARFGYFSAALAPHLIGFELTVLGIYQNPINYSVVALLSIA 540

QY 541 GVLVSGFLANIQEMPPIKIIISFTFQKCYSEILVNVFEYGLNFTCGSSNVSVTTPMC 600
 DB 541 GVLVSGFLANIQEMPPIKIIISFTFQKCYSEILVNVFEYGLNFTCGSSNVSVTTPMC 600
 QY 601 AFTQGIQFIKTCPCGATSRFTMFLIXSTIPALYIIGIYVFKIRHLISR 651
 DB 601 AFTQGIQFIKTCPCGATSRFTMFLIXSTIPALYIIGIYVFKIRHLISR 651
 RESULT 2
 ID AAE13290 standard; Protein: 651 AA.
 AC AAE13290;
 DT 12-FEB-2002 (first entry)
 DE Human strossterolaemia susceptibility gene (SSG) protein.
 KW Human; strossterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.
 OS Homo sapiens.
 PN WO200179272-A2.
 PD 25-OCT-2001.
 PE 18-APR-2001; 2001WO-0512758.
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 PA (TUL-) TULARIK INC.
 PI Tian H, Schultz J, Shan B;
 DR WPI: 2002-017598/02.
 DR N-PSDB: AAD22009.
 PT Novel strossterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 PS Claim 19; Fig 8; 105pp; English.
 XX
 CC The invention relates to an isolated strossterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including strossterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and strossterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing strossterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG protein. Human SSG is located on chromosome
 CC 2p21.
 XX
 SQ Sequence 651 AA:
 Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSIGILHASYSVSHRVRPMWMDITSC 60
Db 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSIGILHASYSVSHRVRPMWMDITSC 60
QY 61 RQOWTROIILKDVSLVYESGQIMCILLGSSGSKTLLDAMSGRLGRAGTFLGEVYVNGRAL 120
Db 61 RQOWTROIILKDVSLVYESGQIMCILLGSSGSKTLLDAMSGRLGRAGTFLGEVYVNGRAL 120
QY 121 RREQFODCFSYVLSQSTLLSLVRETLYTALLAIRNGNPGSFQKVEAVMAELSLSHV 180
Db 121 RREQFODCFSYVLSQSTLLSLVRETLYTALLAIRNGNPGSFQKVEAVMAELSLSHV 180
QY 181 ADRLIGNYSLGISTGERRRVSTIAAQLDOPKVMLEPDEPTGIDCKTANOIYVLLVELAR 240
Db 181 ADRLIGNYSLGISTGERRRVSTIAAQLDOPKVMLEPDEPTGIDCKTANOIYVLLVELAR 240
QY 241 RNRIVVLTIHOPRSELFQFLDKITAILSFGEILFCGTPAEMLDFNDCGYPCEHSNPFDF 300
Db 241 RNRIVVLTIHOPRSELFQFLDKITAILSFGEILFCGTPAEMLDFNDCGYPCEHSNPFDF 300
QY 301 YMDLTSVDIOSKERELETISKRVOMIESAVKSAICHKTLKNIERMKHLKTLPMVPFKTKD 360
Db 301 YMDLTSVDIOSKERELETISKRVOMIESAVKSAICHKTLKNIERMKHLKTLPMVPFKTKD 360
QY 361 SPGVFSKLGVLRRVTRNLRNKLAVITRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420
Db 361 SPGVFSKLGVLRRVTRNLRNKLAVITRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420
QY 421 GLIXYQVAGAPRYGMLNANVLFVLRVAVSQESQDGLYQKWMQMLAYALHVPFVYATM 480
Db 421 GLIXYQVAGAPRYGMLNANVLFVLRVAVSQESQDGLYQKWMQMLAYALHVPFVYATM 480
QY 481 IFSSVCYMTLGLHPEVARFGYFSALLAPHLIGEFITLVILGIVONPNINYSVALLSIA 540
Db 481 IFSSVCYMTLGLHPEVARFGYFSALLAPHLIGEFITLVILGIVONPNINYSVALLSIA 540
QY 541 GVLVSGFLRNIDQEMDIPFKIISYFTFOKYSEILVNEFYGLNFTCGSSNVSATVTPMPC 600
Db 541 GVLVSGFLRNIDQEMDIPFKIISYFTFOKYSEILVNEFYGLNFTCGSSNVSATVTPMPC 600
QY 601 APTOGIOFTEKTCRGATSRFTMNFLLISFTIPALVILGIVKIRDHLSIR 651
Db 601 APTOGIOFTEKTCRGATSRFTMNFLLISFTIPALVILGIVKIRDHLSIR 651

RESULT 3
AA096992
ID AA096992 standard; Protein; 651 AA.
AC AA096992;
DT 30-JUL-2002 (first entry)
DE Human ABCG5 mutant E146Q protein sequence.
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
  arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
  mutant; mouse.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 146 /note= "Wild-type Glu substituted by Gln"
FT
XX WO200227016-A2.
XX 04-APR-2002.
XX 25-SEP-2001; 2001WO-US29859.

```

```

PR 25-SEP-2000; 2000US-235268P.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (PATEL S B.
PA (DEAN/) DEAN M.
XX Patel SB, Dean M;
XX WPI: 2002-416483/44.
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX Claim 12; Page -: 66pp; English.
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant E146Q protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild type human ABCG5 protein (AA096984) given on pages 35-36
CC of the specification.
XX Sequence 651 AA:
SQ
Query Match 99.98; Score 3323; DB 23; Length 651;
Best Local Similarity 99.88; Pred. No. 0;
Matches 650; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSIGILHASYSVSHRVRPMWMDITSC 60
Db 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSIGILHASYSVSHRVRPMWMDITSC 60
QY 61 RQOWTROIILKDVSLVYESGQIMCILLGSSGSKTLLDAMSGRLGRAGTFLGEVYVNGRAL 120
Db 61 RQOWTROIILKDVSLVYESGQIMCILLGSSGSKTLLDAMSGRLGRAGTFLGEVYVNGRAL 120
QY 121 RREQFODCFSYVLSQSTLLSLVRETLYTALLAIRNGNPGSFQKVEAVMAELSLSHV 180
Db 121 RREQFODCFSYVLSQSTLLSLVRETLYTALLAIRNGNPGSFQKVEAVMAELSLSHV 180
QY 181 ADRLIGNYSLGISTGERRRVSTIAAQLDOPKVMLEPDEPTGIDCKTANOIYVLLVELAR 240
Db 181 ADRLIGNYSLGISTGERRRVSTIAAQLDOPKVMLEPDEPTGIDCKTANOIYVLLVELAR 240
QY 241 RNRIVVLTIHOPRSELFQFLDKITAILSFGEILFCGTPAEMLDFNDCGYPCEHSNPFDF 300
Db 241 RNRIVVLTIHOPRSELFQFLDKITAILSFGEILFCGTPAEMLDFNDCGYPCEHSNPFDF 300
QY 301 YMDLTSVDIOSKERELETISKRVOMIESAVKSAICHKTLKNIERMKHLKTLPMVPFKTKD 360
Db 301 YMDLTSVDIOSKERELETISKRVOMIESAVKSAICHKTLKNIERMKHLKTLPMVPFKTKD 360
QY 361 SPGVFSKLGVLRRVTRNLRNKLAVITRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420

```

Db 361 SPGVSKGLVLLRRVTRNLVYNNKLAIVITRLQNLIMGLFLEFVLRAVSNTLKGAIDRV 420
 QY 421 GLLYQVFGATPYTGMNLAVNLEFVPLRAVSDQSDGLYQKQOMMLAYALHVLPSVATM 480
 Db 421 GLLYQVFGATPYTGMNLAVNLEFVPLRAVSDQSDGLYQKQOMMLAYALHVLPSVATM 480
 QY 481 IFSSVCYWTGLAHPVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 Db 481 IFSSVCYWTGLAHPVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSEILVYNEFGLNFTCGSSNVSTYTNMC 600
 Db 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSEILVYNEFGLNFTCGSSNVSTYTNMC 600
 QY 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
 Db 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 4
 AA096989 standard; Protein: 651 AA.
 AA096989;
 AC AA096989;
 XX 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419H protein sequence.
 XX Human ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; muten.
 XX Homo sapiens.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 419 /note= "Wild-type Arg substituted by His"
 FT MO200227016-A2.
 XX 04-APR-2002.
 PD 25-SEP-2001; 2001WO-US29859.
 PF 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX Claim 9; Page -: 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal

CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419H protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.

SO Sequence 651 AA;
 QY Query Match 99.8%; Score 3321; DB 23; Length 651;
 Db Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLVNRGSSSLGAPAPATPEPHSIGILHASYSVSHRVRPMDITSC 60
 Db 1 MGDLSLTPGSGMGLVNRGSSSLGAPAPATPEPHSIGILHASYSVSHRVRPMDITSC 60
 QY 61 RQWTRQILKDVSLVYESGQIMCIIIGSSGKTTLLDAMSGRLAGFTLGEVYNGRAL 120
 Db 61 RQWTRQILKDVSLVYESGQIMCIIIGSSGKTTLLDAMSGRLAGFTLGEVYNGRAL 120
 QY 121 RREOFODCFSTYVLOSSTLSSLVRETLHYTALLAIRGNDSFQKKEAVMAELSLSHV 180
 Db 121 RREOFODCFSTYVLOSSTLSSLVRETLHYTALLAIRGNDSFQKKEAVMAELSLSHV 180
 QY 181 ADRLIGNSLIGISIGERRRYSIAQQLQDPKVMLEFDEPTGLQCMRANOVLVLELAR 240
 Db 181 ADRLIGNSLIGISIGERRRYSIAQQLQDPKVMLEFDEPTGLQCMRANOVLVLELAR 240
 QY 241 RNRIVYLTIHQPSSELPQIDFKIALISFGLIFCPTPEMDLDFNDGCPCEPSNPDEF 300
 Db 241 RNRIVYLTIHQPSSELPQIDFKIALISFGLIFCPTPEMDLDFNDGCPCEPSNPDEF 300
 QY 301 YMDLTSVDTQSKREIEFISKRVQMIESAAYKKAICHKTKLNIEMKHLKTLPMVPFKTKD 360
 Db 301 YMDLTSVDTQSKREIEFISKRVQMIESAAYKKAICHKTKLNIEMKHLKTLPMVPFKTKD 360
 QY 361 SPGVSKGLVLLRRVTRNLVYNNKLAIVITRLQNLIMGLFLEFVLRAVSNTLKGAIDRV 420
 Db 361 SPGVSKGLVLLRRVTRNLVYNNKLAIVITRLQNLIMGLFLEFVLRAVSNTLKGAIDRV 420
 QY 421 GLLYQVFGATPYTGMNLAVNLEFVPLRAVSDQSDGLYQKQOMMLAYALHVLPSVATM 480
 Db 421 GLLYQVFGATPYTGMNLAVNLEFVPLRAVSDQSDGLYQKQOMMLAYALHVLPSVATM 480
 QY 481 IFSSVCYWTGLAHPVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 Db 481 IFSSVCYWTGLAHPVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSEILVYNEFGLNFTCGSSNVSTYTNMC 600
 Db 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSEILVYNEFGLNFTCGSSNVSTYTNMC 600
 QY 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
 Db 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 5
 AA096990 standard; Protein: 651 AA.
 AA096990;
 AC AA096990;
 XX 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419H protein sequence.
 XX Human ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; muten.
 XX Homo sapiens.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 419 /note= "Wild-type Arg substituted by His"
 FT MO200227016-A2.
 XX 04-APR-2002.
 PD 25-SEP-2001; 2001WO-US29859.
 PF 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX Claim 9; Page -: 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal

DE Human ABCG5 mutant R389H protein sequence.
XX
KW Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
mutant; muten.
XX
OS Homo sapiens.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 389 /note- "Wild-type Arg substituted by His"
XX
XX WO200227016-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX
XX 25-SEP-2000; 2000US-235268P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (PATE/) PATEL S B.
XX (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX
XX WPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
XX
XX Claim 7: Page -; 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant R389H protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
CC of the specification.
XX
XX Sequence 651 AA:
SQ

Query Match 99.8%; Score 3321; DB 23; Length 651.
Best Local Similarity 99.8%; Pred. No. 0;
Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGDLSSLPGGSMGLQVNRGSSSLGAPATAPAPPHSGILHASYSHRYRPMWDTISC 60
DB 1 MGDLSSLPGGSMGLQVNRGSSSLGAPATAPAPPHSGILHASYSHRYRPMWDTISC 60
QY 61 RQMTROLKLVSLYVESGQIMCTILGSSGKTTLLDMSGRLGAGTFLGEVYNGRAL 120
DB 61 RQMTROLKLVSLYVESGQIMCTILGSSGKTTLLDMSGRLGAGTFLGEVYNGRAL 120

QY 121 RREOFODCFSYVLOSPTLLSLVRETLYHTALLAIRGNPGSFQKKVAVNAELSLSHV 180
DB 121 RREOFODCFSYVLOSPTLLSLVRETLYHTALLAIRGNPGSFQKKVAVNAELSLSHV 180
QY 181 ADRLIGNYSILGISTGERRRVSTAAQLDOPKVMLEFDEPTGIDCQTANOIVLVLELAR 240
DB 181 ADRLIGNYSILGISTGERRRVSTAAQLDOPKVMLEFDEPTGIDCQTANOIVLVLELAR 240
QY 241 RNRIVVLTTHORSESLFQLPDKTALISFEBLIFCGTPAEMLDFENDCGYPCPEHNPDE 300
DB 241 RNRIVVLTTHORSESLFQLPDKTALISFEBLIFCGTPAEMLDFENDCGYPCPEHNPDE 300
QY 301 YMDLTSVDPQSKERELETSEKRVOMIESAKKSKICHRTKNIERMKHLTLPVPRKTD 360
DB 301 YMDLTSVDPQSKERELETSEKRVOMIESAKKSKICHRTKNIERMKHLTLPVPRKTD 360
QY 361 SPGVESKLGVLRLRYTRNLVRNKLAVITRLQNLINGLFLFVLRVRSNVLKGAIQDRV 420
DB 361 SPGVESKLGVLRLRYTRNLVRNKLAVITRLQNLINGLFLFVLRVRSNVLKGAIQDRV 420
QY 421 GLLYOFVGATPYTGMLNANVLPVLRVAVSDQSDGLYQKQMMALAYALHVLFFSVATM 480
DB 421 GLLYOFVGATPYTGMLNANVLPVLRVAVSDQSDGLYQKQMMALAYALHVLFFSVATM 480
QY 481 IFSSVCYMTLGLHPEVARGVFSALLLAPHLIGELTLVLGIIVONPRIVNSVVALLSIA 540
DB 481 IFSSVCYMTLGLHPEVARGVFSALLLAPHLIGELTLVLGIIVONPRIVNSVVALLSIA 540
QY 541 GVLVSGSFLRNIOENPIPEKIISYTFQKCEIIVNVNFGYGLNFTCGSSNVSVTNPMC 600
DB 541 GVLVSGSFLRNIOENPIPEKIISYTFQKCEIIVNVNFGYGLNFTCGSSNVSVTNPMC 600
QY 601 AFTGCIQFIETKCPGATSRFTNMFLLVSLFALVILGIVPKIRIDHLISR 651
DB 601 AFTGCIQFIETKCPGATSRFTNMFLLVSLFALVILGIVPKIRIDHLISR 651

RESULT 6
AAU96993 standard; Protein; 651 AA.
AAU96993;
30-JUL-2002 (first entry)
Human ABCG5 mutant R419p protein sequence.
Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
mutant; muten.
Homo sapiens.
Synthetic.
Key Location/Qualifiers
FT Misc-difference 419 /note- "Wild-type Arg substituted by Pro"
XX
XX WO200227016-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX
XX 25-SEP-2000; 2000US-235268P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (PATE/) PATEL S B.
XX (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX

CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289) and
CC shown in figure 7 of the specification.

XX Sequence 652 AA:

Query Match 82.6%; Score 2748.5; DB 23; Length 652;
Best Local Similarity 80.4%; Pred. No. 1.1e-280;
Matches 524; Conservative 64; Mismatches 63; Indels 1; Gaps 1;

```
OY 1 MGDLSSLPFGSGMGQVNRGSSSLEGAPATAPPE-HSLGIIMASYSVSHRVRPMWDTTS 59
DB 1 MGELPFLSPGARGPHNRGSLSSLEQASVTGETEARHSIGVLHVSYSVSNRGPMWNITKS 60
OY 60 CROQWTRQIILNDVSLYVESGOIMCIISSSGSGKTTLDAMSGRGLRACGTFGEVYVNGRA 119
DB 61 CQKMDROIILNDVSLYIESGOIMCIISSSGSKTTLDNISGRILKRTGLEEVVNGCE 120
OY 120 LRROFQDCFSYVLOSDFLSSLVRETLRYTALAIIRGNGSPFOKREAVVMAELSLSH 179
DB 121 LRROFQDCFSYVLOSDFLSSLVRETLRYTALAIIRGNGSPFOKREAVVMAELSLSH 180
OY 180 VADRILGNYSLGISTGERRRVSTAQIILDPKVMLEPDEPTGIDCKMANOIVLVLELA 239
DB 181 VADQMGISYNFGISSGERRRVSIAQIILDPKVMLEPDEPTGIDCKMANOIVLVLELA 240
OY 240 RNRRIIVLTIHQPSSELPQFDKTAIISFGELIFCGTPAEMULDEFNDGYPCEPHSNPFD 299
DB 241 RNRRIIVLTIHQPSSELPQFDKTAIILYGLVFCGTPAEMULDEFNDGYPCEPHSNPFD 300
OY 300 FYMDLTSVDQSKEREIETSKRVOMIESAYKKSACHTLKNTERMKHLKTLPMVPFRTK 359
DB 301 FYMDLTSVDQSKEREIETSKRVOMIECAFESDIHYKLENIERARLYKTLPMVPFRTK 360
OY 360 DSPGVFSKIGVLLRRVTRNLVNRKLAIVTTRLLQNLIMCLFLLFVLRLVRSNVLKCAIDR 419
DB 361 DPGMFGKIGVLLRRVTRNLVNRKLAIVTTRLLQNLIMCLFLLFVLRLVRSNVLKCAIDR 420
OY 420 VGLLYQFVGATPYTGMLNAVILFVLRVAVSDQSDGILKHKOMLAIVLVLPFSVAT 479
DB 421 VGLLYQFVGATPYTGMLNAVILFVLRVAVSDQSDGILKHKOMLAIVLVLPFSVAT 480
OY 480 MIFSSVCYTWLGLHPEVARFGYSFSAIILPHLIGELTLVLGIVQNPINVSVALLSI 539
DB 481 VIFSSVCYTWLGLHPEVARFGYSFSAIILPHLIGELTLVLGIVQNPINVSVALLSI 540
OY 540 AGVLVSGEFLINIDEMPIPKIISYFTQKCYCELLVNEVEYGLNFTCGSSNVSTTPM 599
DB 541 SGLLISGSEFLINIDEMPIPKIISYFTQKCYCELLVNEVEYGLNFTCGSSNVSTTPM 600
OY 600 CAFTQGIQFIETPCGATSRFTMFLIXSFIPALVILGIYVFKIRDLISR 651
DB 601 CAFTQGIQFIETPCGATSRFTMFLIXSFIPALVILGIYVFKIRDLISR 652
```

RESULT 8

AAE13289 standard; Protein; 652 AA.

XX AAE13289;

XX 12-FEB-2002 (first entry)

XX Mouse sitosterolemia susceptibility gene (SSG) protein.

XX Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis;
XX sterol-related disorder; hyperlipidaemia; hypercholesterolemia;
XX gall stone; coronary heart disease; cardiovascular disease; arthritis;
XX xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

OS Mus sp.

XX WO200179272-A2.

PD 25-OCT-2001.

XX 18-APR-2001; 2001WO-US12758.

XX 18-APR-2000; 2000US-198465P.

XX 15-MAY-2000; 2000US-204234P.

XX (TULIA-) TULARK INC.

XX Tian H, Schultz J, Shan B;

XX WPI; 2002-017598/02.

XX N-PSDB; AAD22008.

PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder

PS Claim 19; Fig 7; 105pp; English.

CC The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG protein. Mouse SSG is located on
CC chromosome 17.

XX Sequence 652 AA:

Query Match 82.5%; Score 2744.5; DB 23; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.9e-280;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

```
OY 1 MGDLSSLPFGSGMGQVNRGSSSLEGAPATAPPE-HSLGIIMASYSVSHRVRPMWDTTS 59
DB 1 MGELPFLSPGARGPHNRGSLSSLEQASVTGETEARHSIGVLHVSYSVSNRGPMWNITKS 60
OY 60 CROQWTRQIILNDVSLYVESGOIMCIISSSGSGKTTLDAMSGRGLRACGTFGEVYVNGRA 119
DB 61 CQKMDROIILNDVSLYIESGOIMCIISSSGSKTTLDNISGRILKRTGLEEVVNGCE 120
OY 120 LRROFQDCFSYVLOSDFLSSLVRETLRYTALAIIRGNGSPFOKREAVVMAELSLSH 179
DB 121 LRROFQDCFSYVLOSDFLSSLVRETLRYTALAIIRGNGSPFOKREAVVMAELSLSH 180
OY 180 VADRILGNYSLGISTGERRRVSTAQIILDPKVMLEPDEPTGIDCKMANOIVLVLELA 239
DB 181 VADQMGISYNFGISSGERRRVSIAQIILDPKVMLEPDEPTGIDCKMANOIVLVLELA 240
OY 240 RNRRIIVLTIHQPSSELPQFDKTAIISFGELIFCGTPAEMULDEFNDGYPCEPHSNPFD 299
DB 241 RNRRIIVLTIHQPSSELPQFDKTAIILYGLVFCGTPAEMULDEFNDGYPCEPHSNPFD 300
OY 300 FYMDLTSVDQSKEREIETSKRVOMIESAYKKSACHTLKNTERMKHLKTLPMVPFRTK 359
DB 301 FYMDLTSVDQSKEREIETSKRVOMIECAFESDIHYKLENIERARLYKTLPMVPFRTK 360
OY 360 DSPGVFSKIGVLLRRVTRNLVNRKLAIVTTRLLQNLIMCLFLLFVLRLVRSNVLKCAIDR 419
DB 361 DPGMFGKIGVLLRRVTRNLVNRKLAIVTTRLLQNLIMCLFLLFVLRLVRSNVLKCAIDR 420
```

OY 420 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALHVPESVAT 479
 DB 421 VGLLYQVGVATPYTGMLNANVLPVLRVAVSDQSDGLYHMKOMMLAYALHVPESVAT 480
 OY 480 MIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIVONPNIVNSVVALLSI 539
 DB 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIVONPNIVNSVVALLSI 540
 OY 540 AGVIVGSGFLRNIOEMPIPKIISYFFQKCYSEILVNEFYGLNFTCGSSNVSVTTNPM 599
 DB 541 SGLLIGSGFLRNIOEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNTSMNLNPM 600
 OY 600 CATTGCGIOTFKTCPCGATSRFTNMFLLYGFIPALVILGIVKIRIDHLISR 651
 DB 601 CATTGCGIOTFKTCPCGATSRFTNMFLLYGFIPALVILGIVKIRIDHLISR 652

RESULT 9

AAE13308 standard; Protein; 652 AA.

AAE13308;

12-FEB-2002 (first entry)

Mouse sitosterolemia susceptibility gene (SSG) protein variant #1.

Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis; muten;
 KM sterol-related disorder; hyperlipidemia; hypercholesterolemia; mutant;
 KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KM xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

Mus sp.
 OS synthetic.

Location/Qualifiers
 Misc-difference 17 /note- "Wild type ile substituted with leu"

W0200179272-A2.

25-OCT-2001.

18-APR-2001; 2001WO-US12758.

18-APR-2000; 2000US-198465P.

15-MAY-2000; 2000US-204234P.

(TDLA-) TULARIK INC.

Tian H, Schultz J, Shan B;

WPI; 2002-017598/02.

Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder

Disclosure; Page -: 105pp; English.

CC The invention relates to an isolated sitosterolemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number

CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Ile17 with Leu.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289) and
 CC shown in figure 7 of the specification.

Sequence 652 AA;

Query Match 82.5%; Score 2742.5; DB 23; Length 652;
 Best Local Similarity 80.2%; Pred. No. 4,76-280;
 Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

OY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPPP-HSLGILNASYSVSHRPMWDITS 59
 DB 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPPP-HSLGILNASYSVSHRPMWDITS 59
 OY 60 CROQWROIILKDVSLYVESGOIMCILLSSGSGKTLIDAMSGRLGRAGTFLGEVYVNGRA 119
 DB 61 CROQWROIILKDVSLYVESGOIMCILLSSGSGKTLIDAMSGRLGRAGTFLGEVYVNGRA 120
 OY 120 LRREPOFDCESYVLSQSDTLSTLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 179
 DB 121 LRREPOFDCESYVLSQSDTLSTLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 180
 OY 180 VADRLINYSLSGISTEERRRVSIAAQLDDPKVMFLFDEPTGGLDCKTANOIVLLVELA 239
 DB 181 VADRLINYSLSGISTEERRRVSIAAQLDDPKVMFLFDEPTGGLDCKTANOIVLLVELA 240
 OY 240 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFNCGCPCHSPSPFD 299
 DB 241 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFNCGCPCHSPSPFD 300
 OY 300 FYMDLTSVDTSQSKREIETSKRVOMISAVKKSIAIKRTLKNIEMRMLKTLPMVPKTK 359
 DB 301 FYMDLTSVDTSQSKREIETSKRVOMISAVKKSIAIKRTLKNIEMRMLKTLPMVPKTK 360
 OY 360 DSRGVESKILVLRVTRNLRNKLAVITRLNLINGLFLFVLVLRVNSVNLKGAIQDR 419
 DB 361 DSRGVESKILVLRVTRNLRNKLAVITRLNLINGLFLFVLVLRVNSVNLKGAIQDR 420
 OY 420 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALHVPESVAT 479
 DB 421 VGLLYQVGVATPYTGMLNANVLPVLRVAVSDQSDGLYHMKOMMLAYALHVPESVAT 480
 OY 480 MIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIVONPNIVNSVVALLSI 539
 DB 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIVONPNIVNSVVALLSI 540
 OY 540 AGVIVGSGFLRNIOEMPIPKIISYFFQKCYSEILVNEFYGLNFTCGSSNVSVTTNPM 599
 DB 541 SGLLIGSGFLRNIOEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNTSMNLNPM 600
 OY 600 CATTGCGIOTFKTCPCGATSRFTNMFLLYGFIPALVILGIVKIRIDHLISR 651
 DB 601 CATTGCGIOTFKTCPCGATSRFTNMFLLYGFIPALVILGIVKIRIDHLISR 652

RESULT 10

AAU96985 standard; Protein; 652 AA.

AAU96985;

30-JUL-2002 (first entry)

Mouse ABCG5 protein.

Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention.

xx Sequence 652 AA;

Query Match 82.0%; Score 2727.5; DB 23; Length 652;
 Best Local Similarity 79.4%; Pred. No. 1.8e-278;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

OY 1 MGDLSLTPGSGMGLQVNRGSSGLEGAPATAPPE-HSLGILHASYSHRYRPMWDITS 59
 DB 1 MGLPFLSPGAPPHNNRGSSGLEEGSVTGEARHSLGLVNFVSFVSNRPMWNKIS 60
 OY 60 RQOQWTRQILKDVSLVSGQIMCIGSSGSKTTILDAMSGRLGRAGTFGEYVNGRA 119
 DB 61 CQOQWTRQILKDVSLVSGQIMCIGSSGSKTTILDAMSGRLGRAGTFGEYVNGRA 120
 OY 120 LRREPODCFSYVLQSDTLSSLVRETLYTALALIRGNPGSFQKVEAVMELSLSH 179
 DB 121 LRROFOPCSVYLLQSDVFLSSLVRETLYTALALIRGNPGSFQKVEAVMELSLSH 180
 OY 180 VADLLIGNVSLGISTGRRRVSTAALQLOPKVMLEPEPTTGDCMTANQIVLLVLA 239
 DB 181 VADMIGNVNGSLSSGERRRVSTAALQLOPKVMLEPEPTTGDCMTANQIVLLVLA 240
 OY 240 RRRNIVYLTIHQPSSELEFQLEKIAISFGELIFCGPAPMELDFNDCGCPESHNFED 299
 DB 241 RRRNIVYLTIHQPSSELEFQLEKIAISFGELIFCGPAPMELDFNDCGCPESHNFED 300
 OY 300 FYMDLTVSDYOSKRELEFETSKRVOMIESAYKSAICHKTKLNIEEMKHLKTLPMVPFKTK 359
 DB 301 FYMDLTVSDYOSKRELEFETSKRVOMIESAYKSAICHKTKLNIEEMKHLKTLPMVPFKTK 360
 OY 360 DSPGVESKLGLVLRVRNRLVNRKLAVTRLLQNLIMLELFFVLRPSNVLKGATIDR 419
 DB 361 NPQGFCKLGLVLRVRNRLVNRKLAVTRLLQNLIMLELFFVLRPSNVLKGATIDR 420
 OY 420 VGLLYQFVGATPYTGMNAVNLFPVLRAVSDQSDGLYOKOMMLAVLALVLPFSYAT 479
 DB 421 VGLLYQFVGATPYTGMNAVNLFPVLRAVSDQSDGLYOKOMMLAVLALVLPFSYAT 480
 OY 480 MIFSSVCYWTGLHPEVARFGYSALALAPHLIGEFLLTVLLGIVQNPINVSVALLSI 539
 DB 481 VIFSSVCYWTGLHPEVARFGYSALALAPHLIGEFLLTVLLGIVQNPINVSVALLSI 540
 OY 540 AGVAVSGSEFLNIQEMPIPKIISYFTFOKCSLELVNVEYGNATFCGSSNVSYTPPM 599
 DB 541 SGLIIGSEFLNIQEMPIPKIISYFTFOKCSLELVNVEYGNATFCGSSNVSYTPPM 600
 OY 600 CAFTQGIQFLEKTCGATSRFTMFLILSYFIPALVILGIYVFKIRHLISR 651
 DB 601 CSMQGIQFLEKTCGATSRFTMFLILSYFIPALVILGIYVFKIRHLISR 652

RESULT 12

AA096991 ID AA096991 standard; Protein: 408 AA.

AC AA096991;

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R408x protein sequence.
 XX
 KM Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutin.
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 408 /note="Wild-type protein truncated at this position"
 FT
 XX
 PM W0200227016-A2.
 PD
 XX 04-APR-2002.
 XX
 PF 25-SEP-2001: 2001WO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS
 XX Claim 10; Page -: 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal, where an increase or decrease in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or decreases ABCG5 activity respectively. The cell culture or mammal comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The ABCG5 biological activity, or level of ABCG5 mRNA, or level of the polypeptide in a cell culture or mammal is also compared with that of a second cell culture or mammal comprising a wild type ABCG5 polypeptide. Stimulation of ABCG5 activity is useful for treating or preventing hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's disease. The method of the invention is useful for increasing cholesterol excretion and/or decreasing cholesterol adsorption. The present amino acid sequence represents the human ABCG5 mutant R408x protein of the invention.

Note: This sequence is not shown in the specification but is derived from the wild-type human ABCG5 protein (AA096984) given on pages 35-36 of the specification.

Query Match 62.6%; Score 2081; DB 23; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.7e-210;
 Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGDLSLTPGSGMGLQVNRGSSGLEGAPATAPPEHSLGILHASYSHRYRPMWDITSC 60
 DB 1 MGLPFLSPGAPPHNNRGSSGLEEGSVTGEARHSLGLVNFVSFVSNRPMWNKIS 60
 OY 61 RQOQWTRQILKDVSLVSGQIMCIGSSGSKTTILDAMSGRLGRAGTFGEYVNGRAL 120
 DB 61 RQOQWTRQILKDVSLVSGQIMCIGSSGSKTTILDAMSGRLGRAGTFGEYVNGRAL 120

QY 121 RRFQDFQDFSVYQSDTLLSLVRETLHTALLAIRNGPSGOKKVEAVMAELSLSHV 180
 DB 121 RRFQDFQDFSVYQSDTLLSLVRETLHTALLAIRNGPSGOKKVEAVMAELSLSHV 180
 QY 181 ADRLIGNYSIGISTGERRRVSIAAOLLOPKVLMDEPTGIDCMANQIVLVLELAR 240
 DB 181 ADRLIGNYSIGISTGERRRVSIAAOLLOPKVLMDEPTGIDCMANQIVLVLELAR 240
 QY 241 RNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGPAEMLDFPNDGYPCEPSNPFDF 300
 DB 241 RNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGPAEMLDFPNDGYPCEPSNPFDF 300
 QY 301 YMDLTSVDQSKRELETETSKRVOMISAKKSAICHKTKLNIEMKHLKTLPMVPPKTKD 360
 DB 301 YMDLTSVDQSKRELETETSKRVOMISAKKSAICHKTKLNIEMKHLKTLPMVPPKTKD 360
 QY 361 SPGVFSKGLVLRRTVRNLRVNRKLAIVITRLQNLINGLFLFELFVLNR 408
 DB 361 SPGVFSKGLVLRRTVRNLRVNRKLAIVITRLQNLINGLFLFELFVLNR 408

RESULT 13

AA096987
 ID AA096987 standard; Protein; 340 AA.

AC AA096987;

DT 30-JUL-2002 (first entry)

DE Hamster ABCG5 protein.

KM Hamster: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

OS *Cricetinae* sp.

PN W0200227016-A2.

PD 04-APR-2002.

PD 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES;
 (PATEL S B.
 (DEAN/) DEAN M.

PI Patel SB, Dean M;

DR WPI: 2002-416483/44.
 N-PSDB; ABR51687.

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases

PS Example 3; Page 46; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the hamster ABCG5 protein of the invention.

SO Sequence 340 AA;

Query Match 41.7%; Score 1387.5; DB 23; Length 340;
 best Local Similarity 78.6%; Pred. No. 2,2e-137;
 Matches 264; Conservative 39; Mismatches 32; Indels 1; Gaps 1;

QY 98 AMSGRGRAGTFLGEYVNGRALRRQDFQDFSVYQSDTLLSLVRETLHTALLAIR 157
 DB 1 AISGRRLRRGTLEGEYVNGRELKRDQDFQDFSVYQSDTLLSLVRETLHTALLAIR 60

QY 158 RGNPGSFQKKVEAVMAELSLSHVADRIGNYSIGISTGERRRVSIAAOLLOPKVLMFD 217
 DB 61 SSSSDFFYDKKVEAVMAELSLSHVADRMIGNYNGGSSGERRRVSIAAOLLOPKVLMFD 120

QY 218 EPTTGIDCMANQIVLVLELARNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGP 277
 DB 121 EPTTGIDCMANQIVLVLELARNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGP 180

QY 278 AEMLDFPNDGYPCEPSNPFDFYMDLTSVDQSKRELETETSKRVOMISAKKSAICHK 337
 DB 181 EEMLDFPNDGYPCEPSNPFDFYMDLTSVDQSKRELETETSKRVOMISAKKSAICHK 240

QY 338 TKNIERMKHLKTLPMVPPKTKDSPGVFSKGLVLRRTVRNLRVNRKLAIVITRLQNLING 397
 DB 241 TKNIERMKHLKTLPMVPPKTKDSPGVFSKGLVLRRTVRNLRVNRKLAIVITRLQNLING 300

QY 398 LFLFFVLVLRNSVNLKGAIQDRVGLLYOFVGAAPT 433
 DB 301 LFLFFVLVLRNSVNLKGAIQDRVGLLYSW-SAPPT 335

RESULT 14

AA096988
 ID AA096988 standard; Protein; 243 AA.

AC AA096988;

DT 30-JUL-2002 (first entry)

DE Human ABCG5 mutant R243X protein sequence.

KM Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutlein.

OS Homo sapiens.
 OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 243 /note= "Wild-type protein truncated at this position"

PN W0200227016-A2.

PD 04-APR-2002.

PD 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES;
 (PATEL S B.
 (DEAN/) DEAN M.

PI Patel SB, Dean M;

XX WPI; 2002-416483/44.
DR
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
PS Claim 13; Page -: 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant R243X protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
CC of the specification..
XX
XX Sequence 243 AA:

Query Match 37 1%; Score 1234; DB 23; Length 243;
Best Local Similarity 100.0%; Pred. No. 2.1e-121;
Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSITPGSGMGQVNRGSSSSLEGAPATAPAPHSIGILHASVSHRYRPMWDTSC 60
DB 1 MGDLSITPGSGMGQVNRGSSSSLEGAPATAPAPHSIGILHASVSHRYRPMWDTSC 60
QY 61 RQOMTRQILKDVSLVESGQIMCJLSSGSGKTTLLDMSGRLAGAGFLEGVYNGRAL 120
DB 61 RQOMTRQILKDVSLVESGQIMCJLSSGSGKTTLLDMSGRLAGAGFLEGVYNGRAL 120
QY 121 RREQDQDFSVYLOSDTLLSITVRETHYALLAIRGNPGSFQKVEAVMAELISHV 180
DB 121 RREQDQDFSVYLOSDTLLSITVRETHYALLAIRGNPGSFQKVEAVMAELISHV 180
QY 181 ADRLIGNYSIGISIGERRRYSIAQQLDDPKVMLFDEPTGLDQMTANOIVLLVELAR 240
DB 181 ADRLIGNYSIGISIGERRRYSIAQQLDDPKVMLFDEPTGLDQMTANOIVLLVELAR 240
QY 241 RNR 243
DB 241 RNR 243

RESULT 15

AA041856 standard; Protein; 144 AA.

AA041856;

08-FEB-2001 (first entry)

Human ORFX ORF1620 polypeptide sequence SEQ ID NO:3240.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

Homo sapiens.

MO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000MO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CUBAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

N-PSDB; AAC76065.

Claim 11; Page 2444; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipariatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antihypoid; antibacterial; antiviral; antifungal; antineumatic;
CC antiinflammatory; antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 144 AA;

Query Match 21.7%; Score 722; DB 21; Length 144;
Best Local Similarity 99.3%; Pred. No. 1.1e-67;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 357 KTKDPSGVSKGLVLRVTRNLTAVRNLAVITRLQNLINGLFLFVLRVRSVVLKAI 416
DB 1 KTKDPSGVSKGLVLRVTRNLTAVRNLAVITRLQNLINGLFLFVLRVRSVVLKAI 60
QY 417 QDRVGLLYQFVGATPYTGMLNVAVLPVLRVRSVSDQDGLYOKQWMLLYALHVLDFSV 476
|||||

Mon Jul 28 09:43:29 2003

us-09-989-981a-6.rag

Page 13

Db 61 ODRVGLYQFVGAPFYTGMLNANVNFVLRVAVSDQESODGLYOKOMMLAYALHVLPEPV 120
OY 477 VATMFSSVCYWTGLHPEVARFG 500
Db 121 VATMFSSVCYWTGLHPEVARIG 144

Search completed: June 11, 2003, 09:12:15
Job time : 43 secs

OM nucleic - nucleic search, using sw model

run on: July 26, 2003, 23:50:29 ; Search time 4351.59 Seconds
(without alignments)
13069.363 Million cell updates/sec

Title: US-09-989-981A-5
Perfact score: 3240

Perfect score: 2340
Sequence: I gtcaggtgagcagcaggy.....aatattcataaactctggy 2340

Scoring table: IDENTITY_NDC .
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: em.esthma:*
2: em.esthma:*
3: em.estln:*
4: em.estlun:*
5: em.estlv:*
6: em.estpl:*
7: em.estro:*
8: em.htc:*
9: gb.estli:*
10: gb.estl2:*
11: gb.htc:*
12: gb.est3:*
13: gb.est4:*
14: gb.est5:*
15: em.estfun:*
16: em.estfun:*
17: em.gss.hum:*
18: em.gss.invt:*
19: em.gss.pln:*
20: em.gss.vrt:*
21: em.gss.vrt:*
22: em.gss.man:*
23: em.gss.mus:*
24: em.gss.pro:*
25: em.gss.prod:*
26: em.gss.phg:*
27: em.gss.vrt1:*
28: gb.gss1:*
29: gb.gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	544	23.2	594	9	AV689089	AV689089 AV689089
2	521	22.3	597	9	AV694671	AV694671 AV694671
3	477	20.4	477	9	AV720911	AV720911 AV720911
4	463.2	19.8	713	10	BB598373	BB598373 BB598373

C	5	429	18.3	432	9	A1033358
C	6	418	17.9	418	9	A1140253
C	7	407.4	17.4	936	10	BF12636
C	8	396.4	16.9	658	14	BY742680
C	9	394.4	16.9	417	9	AV695922
C	10	365	15.6	471	12	BM856449
C	11	360	15.4	360	9	AV660973
C	12	314.4	13.4	794	14	CA316999
C	13	309.8	13.2	393	9	A1597378
C	14	307.2	13.1	424	9	AA656720
C	15	297	12.7	336	14	T93792
C	16	288.2	12.3	764	12	B1246557
C	17	286	12.2	722	10	BB667343
C	18	281.4	12.0	374	9	AA511669
C	19	281	12.0	460	9	AA230884
C	20	266.4	11.4	356	14	T86385
C	21	233.2	10.0	331	14	TH6684
C	22	211	9.0	625	9	AW112016
C	23	207	8.8	339	10	BB869579
C	24	206	8.8	502	9	AA237916
C	25	206	8.8	535	9	AA244605
C	26	203.6	8.7	706	29	AG094162
C	27	199.2	8.5	3623	11	AK004871
C	28	195	8.3	275	9	A1592875
C	29	188.4	8.1	606	14	CD502116
C	30	187	8.0	783	13	BX092855
C	31	179.4	7.7	516	9	AA237183
C	32	178.6	7.6	366	14	T93842
C	33	172.4	7.4	2417	11	AKA050938
C	34	161	6.9	1245	14	CB813243
C	35	147.4	6.3	357	14	CB813243
C	36	147.2	6.3	358	14	CB812866
C	37	142.2	6.1	442	9	AA217272
C	38	141.6	6.1	695	29	AG122753
C	39	139.8	6.0	833	10	BF620684
C	40	135.4	5.8	610	12	BM725726
C	41	135.4	5.8	749	13	BU740584
C	42	125.8	5.4	527	9	A1497110
C	43	125.8	5.4	816	10	BC995453
C	44	125	5.3	665	28	NZ051029
C	45	122	5.2	855	29	BZ637045
						A1033358 OX2ef10.s
						A1140253 qe2ia04.x
						BF12656 6017693070
						BY742680 BY742680
						AV695922 AV695922
						BM856449 K-E-EST0140
						AV660973 AV660973
						CA316999 UI-M-FW0
						A1597378 vj29c06.y
						AA656720 vp95e08.r
						T93792 ye05f01.s1
						B1246567 602958477
						BB667343 BB667343
						AA511669 vj29c06.r
						AA230884 mx81d01.r
						T86384 yd77b08.r1
						TH6685 yd77b08.s1
						AW112016 MC8117 mo
						BB869579 BB869579
						AA237916 mx14e08.r
						AA244605 mx02d10.r
						AG094162 Pan trolg1
						AK004871 Mus muscu
						A1592875 vp95e08.y
						CD502116 CDA54-H04
						BX092855 BX092855
						AA237183 mx17f01.r
						T93842 ye05f01.r1
						AKA050938 Mus muscu
						CD502117 CDA54-H04
						CB813243 AMGNNUC.T
						CB812866 AMGNNUC.T
						AA217272 ms89f07.r
						AG122753 Pan trolg1
						BF620684 HFSMC002
						BM725726 UI-E-EO-
						BU740584 UI-E-EO-
						A1497110 FB60C04.Y
						BG29523 HVSMA002
						NZ051299 silCO006
						BZ637045 OGCA087C

ALIGNMENTS

RESULT	1
AV689089	
LOCUS	594 bp mRNA linear EST 16-JAN-2002
DEFINITION	AV689089 GKC Homo sapiens cDNA clone GCDCDZB07 5', mRNA sequence.
VERSION	AV689089
KEYWORDS	EST
SOURCE	GI:10290952
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 594)
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,

TITLE

by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

MEDLINE

PUBMED

COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)

Db	477	GGNGGTGCTTGTGATCTGATTCCTCAGAAACATACAGAAATGCCCATCTTTTAA	536
RESULT 3			
AV720911/c			
LOCUS	AV720911	477 bp	mRNA linear EST 16-OCT-2000
DEFINITION	AV720911 GLC Homo sapiens cDNA clone GJCETC06 5', mRNA sequence.		
ACCESSION	AV720911		
VERSION	AV720911.1	GI:10818063	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 477)		
AUTHORS	Qian, B., Wu, T., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.		
TITLE	Homo sapiens cDNA GLC clones		
JOURNAL	Unpublished		
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.		
FEATURES			
source	1. 477		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="GJCETC06"		
	/tissue_type="corresponding non cancerous liver tissue"		
	/dev_stage="adult"		
	/lab_host="SOLR"		
	/clone_lib="GLC"		
	/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	148 a	103 c	95 g
ORIGIN			131 t
Query Match	20.4%; Score 477; DB 9; Length 477;		
Best Local Similarity	100.0%; Pred. No. 6.1e-112;		
Matches 477; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Db	1843	GTTCAGGAGCATGATTTCTGCTGGCAGCTCAAAATGTTCTGTGACAACATCAAT	1902
Db	477	GTTCAGGAGCATGATTTCTGCTGGCAGCTCAAAATGTTCTGTGACAACATCAAT	418
Db	1903	GTCGCTTCACATCAAGGAATTCATTCAGAAACCTGCCAGGTGCAACTCTAG	1962
Db	417	GTCGCTTCACATCAAGGAATTCATTCAGAAACCTGCCAGGTGCAACTCTAG	358
Db	1963	ATTCACAAATGAACTTCTGATTTGATTCATTTATCCAGCTCTTGATCTAGGAT	2022
Db	357	ATTCACAAATGAACTTCTGATTTGATTCATTTATCCAGCTCTTGATCTAGGAT	298
Db	2023	AGTTTCTTCAAAATTAAGGATCATCTATTAGCAGTAGTGAAGCCATGCTGGAA	2082
Db	297	AGTTTCTTCAAAATTAAGGATCATCTATTAGCAGTAGTGAAGCCATGCTGGAA	238
Db	2083	ATGAAGTAGAGCTGCGAGCTGATGATGATGCTCTGAACGCTCAAAATGAGAGTCCAT	2142
Db	237	ATGAAGTAGAGCTGCGAGCTGATGATGATGCTCTGAACGCTCAAAATGAGAGTCCAT	178
Db	2143	GATTTCTTCTTGAAGACATCTCAAGTCTTTTAACTTAAGACATCTTGTGCT	2202
Db	177	GATTTCTTCTTGAAGACATCTCAAGTCTTTTAACTTAAGACATCTTGTGCT	118
Db	2203	CTTGATTCAGACAGGCTTGAATGCAATGAGAGTGTATTAGTCCCTCTCTTACA	2262
FEATURES			
source	1. 713		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
Db	117	CTTGATTCAGACAGGCTTGAATGCAATGAGAGTGTATTAGTCCCTCTCTTACA	58
Db	2263	CTTGATTCAGACAGGCTTGAATGCAATGAGAGTGTATTAGTCCCTCTCTTACA	2319
Db	57	CTTGATTCAGACAGGCTTGAATGCAATGAGAGTGTATTAGTCCCTCTCTTACA	1
RESULT 4			
BB598373			
LOCUS	BB598373	713 bp	mRNA linear EST 26-OCT-2001
DEFINITION	BB598373 RIKEN full-length enriched, adult male liver tumor Mus		
ACCESSION	BB598373		
VERSION	BB598373.2	GI:16450340	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 713)		
AUTHORS	Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda, M., Koye, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, T., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.		
TITLE	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)		
JOURNAL	Unpublished		
COMMENT	On Dec 1, 2000 this sequence version replaced gi:11506974. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Mataliki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing, pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000) Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		
FEATURES			
source	1. 713		
	/organism="Mus musculus"		
	/mol_type="mRNA"		

QY 2088 AGTAAGCTGCGGACTGTGATGACTGCTGTGAACGTCCTGAATGAGATGCCATGTAT 2147
 DB 252 AGTAAGCTGCGGACTGTGATGACTGCTGTGAACGTCCTGAATGAGATGCCATGTAT 193
 QY 2148 TCTTCTTGACAGACATCTCAAGCTCTTTTAAACATTAAAGATCCATTGTGCTCTTGG 2207
 DB 192 TCTTCTTGACAGACATCTCAAGCTCTTTTAAACATTAAAGATCCATTGTGCTCTTGG 133
 QY 2208 ATCCAGCAGAGCCCTTGAATGCAATGAAGAGCTTTATAGTCCCTGCTCTTACACTTGC 2267
 DB 132 ATCCAGCAGAGCCCTTGAATGCAATGAAGAGCTTTATAGTCCCTGCTCTTACACTTGC 73
 QY 2268 AGGACATCTGCTTATTTGAAATTTGAGCTGAGAGCCGACCAAGAAATGAATTAATTC 2327
 DB 72 AGGACATCTGCTTATTTGAAATTTGAGCTGAGAGCCGACCAAGAAATGAATTAATTC 13
 QY 2328 ATTAACCTA 2336
 DB 12 ATTAACCTA 4

RESULT 6
 A1140253 418 bp mRNA linear EST 29-OCT-1998
 LOCUS A1140253
 DEFINITION qe21a04.x1 Soares_fetal_lung.NBHL19W Homo sapiens cDNA clone
 IMAGE:1739598 3', mRNA sequence.
 ACCESSION A1140253
 VERSION A1140253.1 GI:3647710
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 TITLE Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1828 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amerisham
 High quality sequence stop: 417.
 Location/Qualifiers
 1. 418
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1739598"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares_fetal_lung.NBHL19W"
 /note="Organ: Lung; Vector: pRTT3D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTACCAATCTGAGTGGAGCGCGCCGCAATTTTCTTTTCTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NBHL19W."

QY 1918 AGAATTCATTCATTGAGAAACCTGCCAGGAGCAACATCTAGATTCACATGAACCT 1977
 DB 418 AGAATTCATTCATTGAGAAACCTGCCAGGAGCAACATCTAGATTCACATGAACCT 359
 QY 1978 TCTGATTTTGTATTCATTATTCAGCTCTTGTCTCATCTAGATTCATTGTTTCAAAAT 2037
 DB 358 TCTGATTTTGTATTCATTATTCAGCTCTTGTCTCATCTAGATTCATTGTTTCAAAAT 299
 QY 2038 AAGGATTCATTCATTGAGAAACCTGCCAGGAGCAACATCTAGATTCACATGAACCT 2097
 DB 298 AAGGATTCATTCATTGAGAAACCTGCCAGGAGCAACATCTAGATTCACATGAACCT 239
 QY 2098 CCGACTGTGATGACTGCTGTGAACGTCGTGAATGAGAGCCATGTATTTCTTCTGA 2157
 DB 238 CCGACTGTGATGACTGCTGTGAACGTCGTGAATGAGAGCCATGTATTTCTTCTGA 179
 QY 2158 CAGACATCTCAAGCTCTTTTAAACATTAAAGATCCATTGTGCTCTTGAATCCAGCAG 2217
 DB 178 CAGACATCTCAAGCTCTTTTAAACATTAAAGATCCATTGTGCTCTTGAATCCAGCAG 119
 QY 2218 GCCGTGATGATGAGAGCTGTTTATAGTCCCTGCTTACAACTTGAGGAGATGT 2277
 DB 118 GCCGTGATGATGAGAGCTGTTTATAGTCCCTGCTTACAACTTGAGGAGATGT 59
 QY 2278 GGTATTTTGAATTTGACTGAGCGGACCAAGAAATGAATTAATTAATTAACCT 2335
 DB 58 GGTATTTTGAATTTGACTGAGCGGACCAAGAAATGAATTAATTAATTAACCT 1

RESULT 7
 BF162656 936 bp mRNA linear EST 30-OCT-2000
 LOCUS BF162656
 DEFINITION 601769307F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3988777 5',
 mRNA sequence.
 ACCESSION BF162656
 VERSION BF162656.1 GI:11042879
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC).
 TITLE Unpublished
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM9197 row: m column: 02
 High quality sequence stop: 686.
 Location/Qualifiers
 1. 936
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3988777"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP_Lu29"
 /note="Organ: Lung; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

FEATURES

source

BASE COUNT 129 a 89 c 80 g 120 t
 ORIGIN
 Query Match 17.9%; Score 418; DB 9; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1e-96;
 Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 245 a 248 c 257 g 185 t 1 others
 ORIGIN

Query Match 17.4% Score 407.4 DB 10 Length 936;
 Best Local Similarity 75.28; Pred. No. 8-2e-94;
 Matches 534; Conservative 0; Mismatches 172; Indels 4; Gaps 2;

```

OY 605 AAGGTGAGAGCCGCTATGGACAGAGTCTGAGCCAGTGGCAGACCGCATGTATGGC 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 AGGTATAGAGGAGTATGACAGAGTGTAGTCCAGTGGCGGACCAATGATGGC 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 665 AACTACAGCTTTGGGGGATTTCCAGAGGTGAGCGCGCGGCTCCATCCAGCCCGAG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 AGCTATATTTTGGGGGATTTCCAGTGGCGAGCGCGCGGAGTTCCATCCAGCCCA 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 725 CTGCTCCAGAGATCTTAAAGTATGCTTTGATGAGCCACCAAGCGCTGAGTCATG 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 CTCTCTTCAAGAGCCCAAGGTCATGATGCTAGATGAGCAACCAAGAGATGAGTCA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 785 ACTGTATACAGATTTGTCTCTCTGAGTGGAGTCCGAGACCGATCCGATTTGGTT 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 ACTGCAATTCATTAATGCTCTCTCTGAGTGGAGTCCGAGAGGACCGAATTTGATT 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 845 CTCACATTCACAGACCCCGCTTCTGAGTCTTTCAGTCTTTCAGCAAAATTCATCC 904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 CTCACATTCACAGACCCCGCTTCTGAGTCTTTCAGTCTTTCAGCAAAATTCATCC 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 905 AGCTTCGAGAGAGCTATTTCTGTGGACAGCCAGCGAATGCTTATTTCTTCATGAC 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 ACTTACGAGAGATGCTGCTGTGGACAGCCAGAGAGATGCTTGGCTTTCATATAC 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 965 TGGCGTACCTTGTCTGAGTCAATCAACCTTTGACTTCTATATGAGCTGACGTCA 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 TGTGTTTACCTTGTCTGAGTCAATCAACCTTTGACTTCTATATGAGCTGACGTCA 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1025 GTGATATCCCAAGAGAGAGAGGGAATATAGAAACCTTCAGAGAGTCCAGATGATGA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 GTGAGACCCCAAGAGAGAGAGGGAATATAGAAACCTTCAGAGAGTCCAGATGATGA 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1085 TCTGCTTCAAGAAATCGACATTTGTATATAAATTGAGG--AATATGAAAGATGA 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 TGTGCTTCAAGAAATCGACATTTGTATATAAATTGAGGAGACATTTGACAGACAGAC 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1143 AACACCTGAAAGTATACCAATGCTTCAAAACCAAGATTCCTGAGTCTTCT 1202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 GATACCTGAAAGTATACCAATGCTTCAAAACCAAGATTCCTGAGTCTTCT 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1203 CTAAACTGGGTGTTCTCTGAGAGAGTACAGAAACT--TGGTGAAATTAAGCTGCG 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 727 GCCAGCTTGGGGTCTCTGAGAGAGGGAATTAAGAAACTCCACGCGGCAATTAAGACGCG 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1261 AGTATTAAGGCTCTCTCTGAGATTCGATCATGAGGCTTGTCTCTCTT 1310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 787 ACGGATTAATGCGCCAGGAGAACTCGTCAAGGCGCTTACACATAT 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 8
 BY742680 658 bp mRNA linear EST 17-DEC-2002
 DEFINITION musculus cDNA full-length enriched, adult male liver tumor Mus
 BY742680
 VERSION BY742680
 KEYWORDS EST.
 SOURCE GI:27168376
 ORGANISM Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.
 1 (bases 1 to 658)
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nishida, I., Osato, N., Saito, R., Suzuki, H., Yamana, K., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D. P., Bull, C., Hume, D. A.,
 Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,
 Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Cordani

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

FEATURES
 SOURCE

L.E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Paven, N. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontus, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynnshaw-Boris, A., Yanoisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Alzawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, R., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 1246851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Alzawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, K., Muramatsu, M. and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 Location/Qualifiers
 1. 658
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="C730040P06"
 /sex="male"
 /tissue_type="liver tumor"
 /dev_stage="adult"

/note- Site1: SalI; Site_2 BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genetic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15',
GAGGAGAGAGGCGCCGCGCAACCTGAGCTTTTTTTTTTTTTTTNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence 15',
GAGGAGAGAGCTTCGAGTAATTAATTAATTAATTCGCCGCCGCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmidscript KS(+/-) after bulk excision from Lambda Pk. I. Tissue used was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge."

Query Match	16.98;	Score 396.4;	DB 14;	Length 658;
Best Local Similarity	78.28;	Pred. No. 4.9e-91;		
Matches 513; Conservative	0;	Mismatches 138;	Indels 5;	Gaps 3

OY		90	TTTGCCTGTCCTGTGGCCATAGGGTGACCTCTCATCTTTGAGCCCCCGGAGGCTCCATGGCTC	149
Dd		2	TTTTCTCTCTGTATTACCATTGGGTGTAGCTGCCCTTTCTGTAGTCAAGAGGAGCCAAGGGCC	61
OY		150	TCCAAgTAAACAGAGGCTCCAGAGCTCCCTGAGGGGGCTTCGCCAOCGCCCGGAGAC	209
Dd		62	CTCACATCAACAGAGAGGGTCTGTGACTCCCTGGACAAGAGTTGGCTCAAGGGCACAGAGG	121
OY		210	CT---CACAGCCTGGGCAATCCCTCCATGGCTCCTTAGAGGTACGCCAACGCCGTAAGCCCT	266
Dd		122	CTCGGCACAGCTTAGGTGTCTCTGCATGTGTCTCTACAGGCTCAGCAACCGTGTGCGGCTT	181
OY		267	GGTGGGACATCATCTTGTCCGGCAGCAGTGGACCGAGGCAGATCTCTCAAGAATGTCTCT	326
Dd		182	GGTGGAACATCAAAATCATGCGCAGCAGAGTGGGACAGGCAATCTCTCAAGATGTCTCT	241
OY		327	TGTACGTGAGAGGGGGGAGATCAATGTGATCTCTGGAAGCTCAAGCTTCGGGAAAACCA	386
Dd		242	TGTATATGGAAGATGTGCCCAGATTTATGTCAATCTTAAAGCATGACGCTCAGGGAAGACCA	301
OY		387	CGTCTGTGAGCGCCATGTCCGGGAGGCTGGGGCGCGCGGGAGACTCTCTGGGGGAGGTGT	446
Dd		302	CGCTGCTGTGAGGCCATCTCCGGGAGGCTCGGGCGGACATGGGAACCTGGAAAGGAGGTGT	361
OY		447	ATGTGAACGGCGGGGGCGGTGGCGCGGGAGCAGTTCCAGAGATGCTTCTCTACGCTCTGC	506
Dd		362	TTTGTGAATGGCTTGGAAGCTGTGCGCAGGGGACCGATTTCCAAAGACTGTGCTTCTCTACGCTCTGC	421
OY		507	AGAGCGAACACCTCTGTGAGAGCCTTCACTGTGCCGAGAGCCTGCACATCAACCGCGCTGC	566
Dd		422	AGAGCGAGCGTTTCTTCTGAGAGCCTCACTGTGGCGAGACTTTCGATACACAGCATGC	481
OY		567	TGGCATTCGCCCGGGCGCAATCCGGGCTCTTCCAGAAAGAGGTGAGAGCGCGTCATGCGAG	626
Dd		482	TGGCCCTCTGCGGAGCTCCCGGAGCTTCTTCAAACAAGAAAGGTAGAGGCACTCATGACAG	541
OY		627	AGCTAGATCTGAGCATGTGTGCAACGACGACATGTGGCAATPAAGCTGTGGGGGGCATTT	686
Dd		542	AGCTGAGACCTGTGAGCGACAGTGGCGGAGCCACCAATGATTTGGCAGCTATAAATTTGGGGG	742
OY		687	CCACGGGTGAGCGGGCGCGGGTCTTCCATTCGACGCCACAGTGTCTCAGATCTTAG	742
Dd		601	CCAACTGGCGAGCGCGCGCGAGT-TCCAATGGAGCCCAACTCTTAGACACCCCAAGG	655

REFERENCES

TITLE

JOURNAL: Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15085-15094 (2001)
MEDLINE: 21625106
PUBMED: 11752456
COMMENT: Contact: Zeguang Han

Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCDWE04"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2
XhoI"
BASE COUNT      70 a      114 c      109 g      124 t
ORIGIN

```

Query Match	Similarity	Score	394.4: DB 9;	Length 417;
Best Local	Similarity	97.9%;	Pred. No. 1.3e-90;	
Matches 411;	Conservative	0;	Mismatches 6;	Indels 3;
			Gaps	14;
QY	1246	GAGAAATAGCTGGCAGTGAATTACGCGCTCTCTTCAGATTCGATTCATGAGGTGTGTCT	1305	
Db	1	GAGAAATAGCTGGCAGTGAATTACGCGCTCTCTTCAGATTCGATTCATGAGGTGTGTCT	60	
OY	1306	CCTTTTCTCTGCTCTGCGGGTCCGACGATGTGCTAAAGGTGCTATCCAGACCGCGT	1365	
Db	61	CCTTTTCTCTGCTCTGCGGGTCCGACGATGTGCTAAAGGTGCTATCCAGACCGCGT	117	
OY	1366	AGGTCTCCCTTTACACGTTTGTGGGGGCGCCGCCCTTACACAGGCATGCTGAACCGCTGTAA	1425	
Db	118	AGGTCTCCCTTTACACGTTTGTGGGGGCGCCGCCCTTACACAGGCATGCTGAACCGCTGTAA	177	
OY	1426	TCTGTTTCCGCTGCTGGAGGCTGTACAGGACAGAGAGTGAAGACGGCCCTTACACGAA	1485	
Db	178	TCTGTTTCCGCTGCTGGAGGCTGTACAGGACAGAGAGTGAAGACGGCCCTTACACGAA	237	
OY	1486	GTGCGACATGATGCTGGCCTATGACATGACATGACAGTCCCTCCCTTACGCGTGTGTGACACAT	1545	
Db	238	GTGCGACATGATGCTGGCCTATGACATGACATGACAGTCCCTCCCTTACGCGTGTGTGACACAT	297	
OY	1546	GATTTTACAGAGTGTGTGTACTACGTGAGAGCGTGTGTGTACTGTCTAAGTTGCCGCAATTTGG	1605	
Db	298	GATTTTACAGAGTGTGTGTACTACGTGAGAGCGTGTGTGTACTGTCTAAGTTGCCGCAATTTGG	357	

Oy 1606 ATATTTTCTGCTGCTCTTGGCCCCCCTTAATGTCATTTCTTAACCTTGTCT 1665
 Db 358 ATATTTTCTGCTGCTCTTGGCCCCCCTTAATGTCATTTCTTGTGCTCT 417
 RESULT 10
 LOCUS BM856449 471 bp mRNA linear EST 06-MAR-2002
 DEFINITION K-EST0140406 S14K402 Homo sapiens cDNA clone S14K402-48-E04 5',
 mRNA sequence.
 ACCESSION BM856449
 VERSION BM856449.1 GI:19212848
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 48 row: E column: 04
 High quality sequence stop: 471.
 FEATURES
 source
 1..471
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S14K402-48-E04"
 /cell_line="K402"
 /lab_host="TOP10P"
 /clone_lib="S14K402"
 /note="Organ: Stomach; Vector: pTZ198P1; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deapped
 with tabacco acid pyrophosphatase (TAP). The deapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli TOP10P by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."
 BASE COUNT 138 a 79 c 101 g 153 t
 ORIGIN
 Query Match 15.6%; Score 365; DB 12; Length 471;
 Best Local Similarity 100.0%; Pred. No. 5.5e-83;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1976 TTTCGATTTGTTTCATTTATTCAGCTTCTTCATCCAGAAATAGTGTTCAAA 2035
 Db 1 TTTCGATTTGTTTCATTTATTCAGCTTCTTCATCCAGAAATAGTGTTCAAA 60
 Oy 2036 ATAAGGATCATCTCATTTAGCAGTACTGAAAGCCATGCGTGGAAATGGAAGTGAAC 2095
 Db 61 ATAAGGATCATCTCATTTAGCAGTACTGAAAGCCATGCGTGGAAATGGAAGTGAAC 120

Oy 2096 TCCGACATGTCAGTACTGCTCTGACCTGTGAATGAGAGTGCATTTCTTTCTT 2155
 Db 121 TCCGACATGTCAGTACTGCTCTGACCTGTGAATGAGAGTGCATTTCTTTCTT 180
 Oy 2156 GACAGACATCTCAAGCTTTTAAACATTAAAGCTCCATTGTGCTCTTGAGATCAAC 2215
 Db 181 GACAGACATCTCAAGCTTTTAAACATTAAAGCTCCATTGTGCTCTTGAGATCAAC 240
 Oy 2216 AGGCTTGAATGCAATGAGAGTGTTPAGCCCTTCTTCAACTTCAGGAGACT 2275
 Db 241 AGGCTTGAATGCAATGAGAGTGTTPAGCCCTTCTTCAACTTCAGGAGACT 300
 Oy 2276 GTGCTTAATTTGAAATTTGTAAGTGGAGCCGACCAAGATGTAATATTTCAATACCT 2335
 Db 301 GTGCTTAATTTGAAATTTGTAAGTGGAGCCGACCAAGATGTAATATTTCAATACCT 360
 Oy 2336 ATGGG 2340
 Db 361 ATGGG 365
 RESULT 11
 LOCUS AV660973 360 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV660973 GLC Homo sapiens cDNA clone GLCNC08 3', mRNA sequence.
 ACCESSION AV660973
 VERSION AV660973.1 GI:9881987
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 360)
 Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 21625106
 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhongjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source
 1..360
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCNC08"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI."
 BASE COUNT 90 a 77 c 70 g 123 t
 ORIGIN
 Query Match 15.4%; Score 360; DB 9; Length 360;
 Best Local Similarity 100.0%; Pred. No. 9.5e-82;
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1577 GCGTTAATCTCTAGAGTTGCCGATTTGATATTTTCTGCTCTTGGCCCCCACC 1636
 Db 1 GCGTTAATCTCTAGAGTTGCCGATTTGATATTTTCTGCTCTTGGCCCCCACC 60

QY 1637 TTAATTGGCAATTTCTAACTCTTGTCTACTTGTGATCGTCCAAATCCAAATATATGTC 1696
 |||||||
 Db 61 TTAATTGGCAATTTCTAACTCTTGTCTACTTGTGATCGTCCAAATCCAAATATATGTC 120
 |||||||
 QY 1697 AACAGCTAGTGGCTCTGCTGTCCATGCGGGGGCTCTGTGGATCGATTCCTCAGA 1756
 |||||||
 Db 121 AACAGCTAGTGGCTCTGCTGTCCATGCGGGGGCTCTGTGGATCGATTCCTCAGA 180
 |||||||
 QY 1757 AACATCAAGAAATGCCCATTCCTTTAAATCATAGTATTTTATTCCTCAAAATAT 1816
 |||||||
 Db 181 AACATCAAGAAATGCCCATTCCTTTAAATCATAGTATTTTATTCCTCAAAATAT 240
 |||||||
 QY 1817 TCCAGTGAATCTTGTACTCATATGATTCATGAGCATGATTTTCACTTGTGACACTCA 1876
 |||||||
 Db 241 TCCAGTGAATCTTGTACTCATATGATTCATGAGCATGATTTTCACTTGTGACACTCA 300
 |||||||
 QY 1877 AATGTTTCTGTACACATCATTCATGTTGCTGCTTCACTCAAGAAATTCATTCATTCAG 1936
 |||||||
 Db 301 AATGTTTCTGTACACATCATTCATGTTGCTGCTTCACTCAAGAAATTCATTCATTCAG 360
 |||||||

RESULT 12
 CA316999 794 bp mRNA linear EST 26-NOV-2002
 LOCUS DEFINITION UI-M-FW0-cbm-a-08-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
 IMAGE: 6811372 5', mRNA sequence.

ACCESSION CA316999
 VERSION CA316999.1 GI:24535123
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 AUTHORS NIH-MGC http://mgl.ncl.nih.gov/.
 TITLE NIH-MGC
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rtm.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMD at:
 http://image.lmd.gov

FEATURES
 source
 Seq primer: pyx-5.
 Location/Qualifiers
 1..794
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6811372"
 /tissue="embryo 13.5,14.5,16.5,17.5dpc"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_id="NIH_BMAP_FW0"
 /note="Organ: Brain; Vector: pyx-Asc; Site: 1: EcoR I;
 Site 2: Not I; The library was constructed according
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

QY 370 AGGCTCCGGAAACACACCGCTCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 429
 |||||||
 Db 87 AGGCTCCGGAAACACACCGCTCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 146
 |||||||
 QY 430 CTTCCTGGGGAGGTGTATGTAAAGCGCCGGCGCTGCCCGGAGACAGTCCAGAGACTG 489
 |||||||
 Db 147 CCTGGAGGGGAGGTGTATGTAAAGCGCTGCCCGGAGACAGTCCAGAGACTG 206
 |||||||
 QY 490 CTTCCTCCAGC----- 500
 |||||||
 Db 207 CTTCCTCCAGCCTCTGACAGGTGGGGCTGTCCCTGAGCCCTGCGGGCTGGGCC 266
 |||||||
 QY 501 -----TCTGAGAGCGACACCCCTGCTGA 524
 |||||||
 Db 267 CTAGCCCGCGGATTTGACAGACCTGATGTCCTTCTCTGAGAGCGAGCTTTTCTGA 326
 |||||||
 QY 525 GCAGCCTCACCCTGGCGGAGAGCTGACACTACACCGCGCTGGCCATCCCGCGGCA 584
 |||||||
 Db 327 GCAGCCTCAGTGTGGCGGAGAGCTGCGATACACAGCGATGCTGGCCCTCTCCAGCT 386
 |||||||
 QY 585 ATCCCGGCTCCTCCAGAGAGAGGTGAGCGCGTATGCGACAGCTGATGAGCCATG 644
 |||||||
 Db 387 CCGCGAGCTCTTACAAACAGAGGTAAAGCGTATGATGACAGAGCTGAGCCGACAGC 446
 |||||||
 QY 645 TGGCAGACCGAGCTGATTTGGCACTACAGTGGGGGAGTTCACAGGCTGAGCGGGCG 704
 |||||||
 Db 447 TGGCAGACCGAGCTGATTTGGCACTACAGTGGGGGAGTTCACAGGCTGAGCGGGCG 506
 |||||||
 QY 705 GGGTCTCCATGCGACCGCAGCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCNA 764
 |||||||
 Db 507 GAGTTCATGCGACCGCAGCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCNA 566
 |||||||
 QY 765 CCACAGGCGCTGAGCTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
 |||||||
 Db 567 CCACAGGCGCTGAGCTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 626
 |||||||
 QY 825 GCAGAGACCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
 |||||||
 Db 627 GCAGAGACCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
 |||||||
 QY 885 TTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
 |||||||
 Db 687 TTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 745
 |||||||
 QY 945 TGCCTGATTT 954
 |||||||
 Db 746 TGCCTGATTT 755
 |||||||

BASE COUNT 156 a 239 c 220 g 176 t 3 others
 ORIGIN
 Query Match 13.4%; Score 314.4; DB 14; Length 794;
 Best Local Similarity 71.8%; Pred. No. 8.5e-70;
 Matches 481; Conservative 0; Mismatches 103; Indels 86; Gaps 2;
 Developing Mouse Nervous System, supported by National
 Institute of Mental Health (NIMH), Hemin Chln, Ph.D.,
 program coordinator."
 370 AGGCTCCGGAAACACACCGCTCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 429
 |||||||
 87 AGGCTCCGGAAACACACCGCTCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 146
 |||||||
 430 CTTCCTGGGGAGGTGTATGTAAAGCGCCGGCGCTGCCCGGAGACAGTCCAGAGACTG 489
 |||||||
 147 CCTGGAGGGGAGGTGTATGTAAAGCGCTGCCCGGAGACAGTCCAGAGACTG 206
 |||||||
 490 CTTCCTCCAGC----- 500
 |||||||
 207 CTTCCTCCAGCCTCTGACAGGTGGGGCTGTCCCTGAGCCCTGCGGGCTGGGCC 266
 |||||||
 501 -----TCTGAGAGCGACACCCCTGCTGA 524
 |||||||
 267 CTAGCCCGCGGATTTGACAGACCTGATGTCCTTCTCTGAGAGCGAGCTTTTCTGA 326
 |||||||
 525 GCAGCCTCACCCTGGCGGAGAGCTGACACTACACCGCGCTGGCCATCCCGCGGCA 584
 |||||||
 327 GCAGCCTCAGTGTGGCGGAGAGCTGCGATACACAGCGATGCTGGCCCTCTCCAGCT 386
 |||||||
 585 ATCCCGGCTCCTCCAGAGAGAGGTGAGCGCGTATGCGACAGCTGATGAGCCATG 644
 |||||||
 387 CCGCGAGCTCTTACAAACAGAGGTAAAGCGTATGATGACAGAGCTGAGCCGACAGC 446
 |||||||
 645 TGGCAGACCGAGCTGATTTGGCACTACAGTGGGGGAGTTCACAGGCTGAGCGGGCG 704
 |||||||
 447 TGGCAGACCGAGCTGATTTGGCACTACAGTGGGGGAGTTCACAGGCTGAGCGGGCG 506
 |||||||
 705 GGGTCTCCATGCGACCGCAGCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCNA 764
 |||||||
 507 GAGTTCATGCGACCGCAGCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCNA 566
 |||||||
 765 CCACAGGCGCTGAGCTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
 |||||||
 567 CCACAGGCGCTGAGCTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 626
 |||||||
 825 GCAGAGACCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
 |||||||
 627 GCAGAGACCGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
 |||||||
 885 TTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
 |||||||
 687 TTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 745
 |||||||
 945 TGCCTGATTT 954
 |||||||
 746 TGCCTGATTT 755
 |||||||

RESULT 13
 A1597378
 LOCUS DEFINITION vj29c06.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
 clone IMAGE:930442 5', mRNA sequence.
 ACCESSION A1597378
 VERSION A1597378.1 GI:4606426
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Mashu-NCI Mouse EST Project 1999
Unpublished
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:535362

This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Seq primer: -40RP from Gldco
High quality sequence stop: 389.

FEATURES

source
1..393
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:930442"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/clone_1lb="Stratagene mouse diaphragm (#937303)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dt. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTGAGATTTTCTTTTCTTTT 3'."

BASE COUNT 90 a 95 c 83 g 125 t
ORIGIN

Query Match 13.2%; Score 309.8; DB 9; Length 393;
Best Local Similarity 86.8%; Pred. No. 9.4e-69;
Matches 341; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 1425 ATCTGTTTCCCGCTGCGAGCTGTGACGACGAGAGAGTCAGCGCTCTACCGAGA 1484
DB 1 ATCTGTTTCCCGCTGCGAGCTGTGACGACGAGAGAGTCAGCGCTCTACCGAGA 60
OY 1485 AGTGGCAGATGATGCTGCTATGACATGACAGTCTCTCCCTGACGCTTGTGCCACCA 1544
DB 61 AGTGGCAGATGATGCTGCTATGACATGACAGTCTCTCCCTGACGCTTGTGCCACCA 120
OY 1545 TGATTTTCAGAGTGTGCTGCTACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
DB 121 TCATTTTCAGAGTGTGCTGCTACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
OY 1605 GATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1664
DB 181 GATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
OY 1665 TACTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
DB 241 TGTGTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
OY 1725 CGGGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1784
DB 301 CTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
OY 1785 AAATCAGCAGTATTTTACATTCACAAAATAT 1817
DB 361 AAATCAGCAGTATTTTACATTCACAAAATAT 393

RESULT 14

AA656720 424 bp mRNA linear EST 04-NOV-1997
LOCUS vps9e08.r1 stratagene mouse diaphragm (#937303) Mus musculus cDNA
DEFINITION clone IMAGE:1092518 5' similar to SW:SCRT_PDR0M P45843 SCARLET
PROTEIN; ; mRNA sequence.
ACCESSION AA656720
VERSION AA656720.1 GI:2592874
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
(bases 1 to 424)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The Mashu-HMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:598750
Seq primer: -28m13 rev1 ET from Amersham.

FEATURES

source
1..424
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1092518"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/clone_1lb="Stratagene mouse diaphragm (#937303)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dt. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTGAGATTTTCTTTTCTTTT 3'."

BASE COUNT 86 a 112 c 113 g 113 t
ORIGIN

Query Match 13.1%; Score 307.2; DB 9; Length 424;
Best Local Similarity 82.8%; Pred. No. 4.6e-68;
Matches 351; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 1186 TTCTCCGAGATTTTCTTAACTGGGAGTCTCTGAGAGAGTGAAGAACTGCT 1245
DB 1 TTCTCCGAGATTTTCTTAACTGGGAGTCTCTGAGAGAGTGAAGAACTGCT 60
OY 1246 GAGAAATTAAGCTGACATGATTAAGCTGCTTCAAGATCTGATGAGTTGTTCT 1305
DB 61 GAGAAATTAAGCTGACATGATTAAGCTGCTTCAAGATCTGATGAGTTGTTCT 120
OY 1306 CCTTTTCTGCTTGTGGGGGCTGCGACCAATGCTTAAAGGCTCTTCCAGAGCGCGT 1365
DB 121 CATTTTCTGCTTGTGGGGGCTGCGACCAATGCTTAAAGGCTCTTCCAGAGCGCGT 180
OY 1366 AGTCTCCTTTACAGATTTGAGGCGCCACCCGCTTACAGAGCATGCTGAAGCTGTGAA 1425
DB 181 GGGGCTGCTCTATAGCTTGTGGGTGCGACCCCTATACCGGCTGCTCAATGCTGTGAA 240
OY 1426 TCTGTTCCCGCTGCTGAGCTGTACAGGACGACGAGAGTGTAGAGCGCTCTACACGAA 1485
DB 241 TCTGTTCCCGCTGCTGAGCGCTGTACAGGACGACGAGAGTGTAGAGCGCTCTATACATA 300

us-09-989-981a-5.rst

Page 11

Query Match	12.7%	Score 297;	DB 14;	Length 336;
Best Local Similarity	96.9%	Pred	No. 1.8e-65;	
Matches 312;	Conservative	0;	Mismatches 8;	Indels 2;
				Gaps 1;

Search completed: July 27, 2003, 08:29:24
Job time : 4356.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:11:47 ; Search time 18-Seconds
(without alignments)
1064.129 Million cell updates/sec

Title: US-09-989-981A-6
Perfect score: 3336
Sequence: 1 MGDLSLTPGSGMGLQVNRG.....PALVILGIVEKINDHLIRL 651

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents.AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682.5	20.5	655	4 US-09-245-808-1	Sequence 1, Appli
2	244	7.3	248	4 US-09-134-001C-3731	Sequence 3731, Ap
3	244	7.3	1280	2 US-08-752-447-2	Sequence 2, Appli
4	244	7.3	1280	4 US-09-316-167-2	Sequence 2, Appli
5	240	7.2	1279	2 US-08-784-649A-2	Sequence 2, Appli
6	240	7.2	1280	2 US-08-583-276-19	Sequence 19, Appli
7	240	7.2	1280	6 5206352-4	Patent No. 5206352
8	239.5	7.2	1684	3 US-08-665-259-25	Sequence 25, Appli
9	239.5	7.2	1684	3 US-08-762-500-25	Sequence 75, Appli
10	239.5	7.2	1704	3 US-09-134-001C-3832	Sequence 3832, Ap
11	230.5	6.9	242	4 US-09-134-001C-3369	Sequence 3369, Ap
12	229	6.9	460	4 US-09-134-001C-3369	Sequence 3641, Ap
13	229	6.9	1307	1 US-08-395-246C-2	Sequence 2, Appli
14	220.5	6.6	1273	4 US-09-134-001C-3641	Sequence 3641, Ap
15	220	6.6	1457	3 US-08-665-259-27	Sequence 27, Appli
16	220	6.6	1457	3 US-08-762-500-27	Sequence 27, Appli
17	218	6.6	1275	3 US-09-120-513-2	Sequence 2, Appli
18	218	6.6	1275	4 US-09-450-105-2	Sequence 2, Appli
19	213	6.4	228	4 US-09-134-001C-4373	Sequence 4373, Ap
20	211	6.3	270	4 US-09-134-001C-5561	Sequence 5561, Ap
21	210	6.3	382	4 US-09-134-001C-3439	Sequence 3439, Ap
22	209.5	6.3	244	4 US-08-919-573-4	Sequence 4, Appli
23	209.5	6.3	341	4 US-09-134-001C-4042	Sequence 4042, Ap
24	208.5	6.3	202	4 US-08-858-207A-332	Sequence 332, Ap
25	208.5	6.3	229	4 US-09-134-001C-4113	Sequence 4113, Ap
26	208.5	6.3	229	4 US-09-627-376-15	Sequence 15, Appli
27	207.5	6.2	246	4 US-09-627-376-15	Sequence 15, Appli

28	206.5	6.2	233	4 US-08-993-825-2	Sequence 2, Appli
29	206.5	6.2	1476	4 US-09-256-703-2	Sequence 2, Appli
30	206.5	6.2	1480	1 US-07-637-621-2	Sequence 2, Appli
31	206.5	6.2	1480	2 US-08-951-912-2	Sequence 2, Appli
32	206.5	6.2	1480	4 US-08-681-838A-2	Sequence 2, Appli
33	206.5	6.2	1480	4 US-08-681-838A-3	Sequence 3, Appli
34	206.5	6.2	1480	4 US-09-174-077-2	Sequence 2, Appli
35	206.5	6.2	1480	6 5240846-5	Patent No. 5240846
36	206	6.2	406	4 US-09-134-001C-3570	Sequence 3570, Ap
37	206	6.2	1479	2 US-08-951-912-4	Sequence 4, Appli
38	206	6.2	1479	4 US-09-174-077-4	Sequence 4, Appli
39	205.5	6.2	1480	4 US-09-425-453A-6	Sequence 6, Appli
40	205.5	6.2	1480	4 US-09-425-453A-12	Sequence 12, Appli
41	205.5	6.2	1480	4 US-09-425-453A-20	Sequence 20, Appli
42	205	6.2	225	4 US-09-134-001C-3215	Sequence 3215, Ap
43	204.5	6.1	1308	2 US-08-996-644-2	Sequence 2, Appli
44	204.5	6.1	1308	3 US-09-352-552-2	Sequence 2, Appli
45	204.5	6.1	1480	1 US-08-136-742A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1	
US-09-245-808-1	
Sequence 1, Application US/09245808	
Patent No. 6313277	
GENERAL INFORMATION:	
APPLICANT: Doyle, L. Austin	
APPLICANT: Abruzzo, Lynne V.	
APPLICANT: Ross, Douglas D.	
TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which	
TITLE OF INVENTION: encodes it	
FILE REFERENCE: Ross UMD conversion	
CURRENT APPLICATION NUMBER: US/09/245, 808	
EARLIER FILING DATE: 1999-02-05	
EARLIER APPLICATION NUMBER: 60/073763	
NUMBER OF SEQ ID NOS: 7	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO. 1	
LENGTH: 655	
TYPE: PR	
ORGANISM: Human MCF-7/AdrVP cells	
US-09-245-808-1	
Query Match	20.5%; Score 682.5; DB 4; Length 655;
Best Local Similarity	29.2%; Pred. No. 1.9e-64;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;	
21 SSSLEGAPATAP---EPHSLGILHASYSVSHRRPMDITSGROQRTQIILKDVSLYE 77	Sequence 2, Appli
13 SGGTNGFPAATASNDLKAFTGAVLSPHNICRYNKLKSGFLPCPKPEKELISNINIMK 72	Sequence 2, Appli
78 SQIMICLSSGSGKTTILDAMSGRLGACFTFGEVYVNGRALRREQFQDCFSYVLSQDT 137	Sequence 2, Appli
73 PG-LMILGPTGGKGLDVLVLAARKDPG-LSGDVLING-APRPANFKNSGVVQDDV 129	Sequence 2, Appli
138 LLSLVVRETLHTALLAIRGNPG--SFQKVAVMAELSLSHVADLLJNYSIGLSTG 196	Sequence 2, Appli
130 VMGTLVREKLSAALRLATYTNHNEKNIRNVIOELGDLVVAOSKVTGPIRQVSG 189	Sequence 2, Appli
197 ERRRVIAQLADPKVMLDEPTTGLDCTANQIVLVLELARRNRIVLTHOPREL 256	Sequence 2, Appli
190 ERKRTSIGMELTDPISLFLDEPTTGLDSTANAVALLLKRMKOGRTIIFSHOPRYSI 249	Sequence 2, Appli
257 FOLFALIALSLGELIFCGTPAEMLDFFNDCCGYPCEHNSPDPFYDLTSVDTQ---SK 312	Sequence 2, Appli
250 FKLFSLTLVLAAGRLMFGPAQEAALGYFSAAGYHCAVYNNPDEFDIINGSTAVANLR 309	Sequence 2, Appli
313 ERE-----IETSKR-----VOMIESAYKKSALCHKT-----LKNTERKKHLKTLPMVPP 356	Sequence 2, Appli
310 EEDFKATELIEPSKQDKPLIEKLALEYVNSSEFYKTKAELHQLSGEKKKKITVFEISY 369	Sequence 2, Appli

QY 357 KTKDSFVSKLGVLRRVTRNLVYKNAVLITRLLQNLIMGLFL--LFFVLRYRSVNLNG 414
DB 370 TT-----SPCHOLRWMSKRSFKMLGNPOASIAQIYTVVGLVIGALVIGLKNDSF----- 421
QY 415 AIDDRVGLLYXOPFVGAFTPYGMNAVLPFVLRAVSQESODGLYKQWMLAYAL-HVLP 473
DB 422 GIONRAGVLF-PLTTNOCSSVSAVELFVEKRLFHEXISQYRVSSEYFLKRLSDLP 480
QY 474 FSVVAMIFSSVCYWTGLHPEVAREGYSAALLAPHLIGEFLLVGLGIONPNVNSV 533
DB 481 KMTLPSTIIFTCLVYFMGLGKPKADAFVMMFTLM--MAYASASSALAIAGQSVSYVA 537
QY 534 VALLSAGV--LVSGFLNIOEMPIPKIISYFFQKYSEILVYNEFYGLNFTCGSSN 591
DB 538 TLMTCFVFMIFSGLLVNLTTIASWLSMLQFSIPRYGFTALQHNFCPCG--- 594
QY 592 VSWTPMCAFTGQIOFIETKCPG 615
DB 595 LNAITGNPCNVA-----TCTG 610

RESULT 2

US-09-134-001C-3731
Sequence 3731, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3731
LENGTH: 248
TYPE: PRP
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3731

Query Match 7.3%; Score 244; DB 4; Length 248;
Best Local Similarity 29.3%; Pred. No. 8.4e-18;
Matches 70; Conservative 53; Mismatches 92; Indels 24; Gaps 9;
QY 67 QILKDVSLVESGQIMCLIGSSGSKTLLDMSG-RLGRAGTFLGEVYNGRALRRE-- 123
DB 20 EVIKIDIKINGEVYTLIGRSGSKTLLIMINALIETEGT---YVNGMTYNTKDK 75
QY 124 ---OFODCFSYLQSDTLLSLVRETLHTALLAIRGNPQSFQKVEAVMAELSLH 179
DB 76 KSOIKVROOSGWFQNTYLPFKHSALENN-MGLITVKKMKKATANEAMMLLAKVGLVH 134
QY 180 VADRLLIGNVSLIGISTGERRRVSIAAQLDOPKVALFDEPTTGDCMTANOIVLVLETA 239
DB 135 VKDQ--RPHALSG--GGQORVATARALANMPKVALFDEPTSDLPDELVNDVAKYKELA 189
QY 240 RNRKIVLTIQPSELFDKTAIISFGELIFCGQPAEMLDFFNDGYPCEHSNPF 298
DB 190 DEGMTVIVTHEMRPAK-EVSNQIAFIHGVIAEGTPE--DIFN--HKTEELQRF 241

RESULT 3

US-08-752-447-2
Sequence 2, Application US/08752447
Patent No. 5994088
GENERAL INFORMATION:
APPLICANT: Mechtner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and

TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: McDonnell Boenhen Hubbert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,447
FILING DATE: 15-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5994088nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-9808
TELEFAX: 312-913-0001
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-447-2

Query Match 7.3%; Score 244; DB 2; Length 1280;
Best Local Similarity 20.7%; Pred. No. 1.5e-16;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;
QY 41 LHASYSVSHRYRPMMDITSCROQTROLKDVSLVESGQIMCLIGSSGSKTLLDMS 100
DB 397 VHFSTPSRKEVK-----ILKGIINTLKVGSGQVVALVNGSGCKSTVQIMQ 441
QY 101 GRLGAGTFLGEVYNGRALR--REOFODCFSYLQSDTLLSLVRETLHYTLLAIR 158
DB 442 RL-YDPTBGMVSVDDGODITINRFLREILIGVSOEPVLFATTIANIRY----- 490
QY 159 GNPQSFQKVAVNAE-----LSLHVADRLLIGNVSLIGISTGERRRVSIAAQLDOP 211
DB 491 GREVTMDIEKAVKAEANAYDFIMKLPKFDLVGERG-AQLSGGQORAIARALVRNP 549
QY 212 KVMLEDEPTTGDCMTANOIVLVLELARRRRIYVLTIHORSELPFLDKAIISFEL 271
DB 550 KILLDEATSLD--TESAVVQVALDKARKKRTTIVIAH--RFATVRNADVIAGFDGVI 606
QY 272 IFCGTPAEMDFPNDGCPCEHSNPFVMDLTSVDTSQKEREIETSK----- 320
DB 607 VEKGNHDELM-----KEKQITKLTVMQAGVELENAADSKSEIDALE 652
QY 321 -----RVQMIESAYKKSALCHRTLNIERMKHLKLPVVF-----RTKDSFG 363
DB 653 MSNDSRSLIRKSTRSRVSGOAHKRLSTKALD--ESIPVSEFRIKMLNITEMPY 710
QY 364 -----VESK-LIGVLLR-----RVTRN-----LYRNKLAVYTR 389
DB 711 FVGVFCAILINGIOPAFALIFSKIIIGFTYIDDPETKRQNSLFSLLFALIGIISFTF 770
QY 390 LLAONLNG-----LELLEFYLVRASNV-----LKGAIQ 417
DB 771 FLGFTFGKAGEILTKRLRYVWFRSMLRQDVSWMHDKNTGALTTRLANDAAGYKAGIG 830
QY 418 DRV-----GLIYQVGAFTPYGMNAVNLPPVL-----RAVSDOE 452
DB 831 SRLAVITQNIANLGTGIISFIYGMQTLTLLAI--VPILAIAGVEMKMFAGALDKK 888

QY 453 SODGL-----YOKOMMLAVALHV-----LPEFSVATM 480
DB 889 ELEGAKIATEAIEENFRVVSITQOKFEHMTAQSLOVRYNSLRKAHIFGTFSETOAM 948
QY 481 IFSSV--CYWTLGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIYONPNIVNSV--- 533
DB 949 MYFSYAGCF-----RFG---AYLVAKHLMSEFDVLLVFSAVVFGAMAVQVSSSF 994
QY 534 -----VALSTAGVLVSGFELRNIOEMPIFKIISYFFQKYCSEILVYNEFYGLNFTCG 588
DB 995 APDYAKAKISAAHIIM-----IIKTPL-----IDYSTEGIMPNTELEG--NVTFG 1038
QY 589 SSNSVTNPMCAFTOGI 606
DB 1039 EVVFNYPTRPDPIDPVLQGL 1056

RESULT 4

US-09-316-167-2
Sequence 2, Application US/09316167
Patent No. 6365357
GENERAL INFORMATION:
APPLICANT: Mechtner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and
TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boenigen Hulbert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/316,167
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/752,447
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: No. 6365357nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-9808
TELEFAX: 312-913-9808
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-316-167-2

Query Match

Best Local Similarity 7.3%; Score 244; DB 4; Length 1280;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

QY 41 LAHSYSYSHVRPMWMDITSCROQWTRQIILNDVSLVYESQIMCILLGSSGSGKTTLLDAMS 100
DB 397 VHSYSRKEVK-----ILKGLNLKTVQSGTVALVNGSGCGKSTYVQLMQ 441
QY 101 GRGGRAGTFLGEYVNGRLR--REOPDCCFSTYLDSDTLLSSLYRETLHYTALLAIR 158
DB 442 -RL--YDPTGAMVSDGQDIRITNVFRLREITGVVSGQEPVLFTFTIAENIRY----- 490

QY 159 GNPSPQKVEAYMAE-----LSLSHVADLLIGNYSIGISTGERRRVSIAAQLDDP 211
DB 491 GRENVTMDIEKAVNEANAYDPMKLPHEFDLVGERG--AQISGQOKORIALARLVNRP 549
QY 212 KVMLEDEPTTGIDCTANOIVLVLELARNRITVLTIIHQPSSELPFQLEDKTAISFGEI 271
DB 550 KILLDEANSALD--TESSEAVQVALDKARKGTTIYIAH--REAVYRNADVLAGDDGYI 606
QY 272 ITCGPBAEMLDFNOCGIPCPEHSPDFEDYMDLTSDVQSKREIETSK----- 320
DB 607 VERGNHDELM-----KEGIIYFLVMTQFAGNEVELENADESKSEIDALE 652
QY 321 -----RVOMIESAYKKSACHTKTKIEEMKHKTLPMPF-----KTKDSFG 363
DB 653 MSSNDRSSLKRRKSTRSRVSGOAHRLSTKEALD--ESTIPVSFWRMKLNLTMPY 710
QY 364 -----VFSK-LGVLLR-----RVFRN-----LVNKLAVTR 389
DB 711 FVYGVFCALINGLQPAFALIESKIIIGVTRIDDETRKONNLESLEFLALGITSFTIF 770
QY 390 LLQNLIMG-----LFLEFVLVRSNV-----LKGAIQ 417
DB 771 FLQGFTEFGAGELIKRLRYWFRSMLKODVSFHDPKNTGALTTRLANDAQAQKAGIG 830
QY 418 DRV-----GLIXQFGATPYTGMLNANVLPVL-----RAVSDOE 452
DB 831 SRLAVITONIANLGTGIIISFIYGMQLTLLAI--VPIIAGVEMKMPAGQALDKDK 888
QY 453 SODGL-----YOKOMMLAVALHV-----LPEFSVATM 480
DB 889 ELEGAKIATEAIEENFRVVSITQOKFEHMTAQSLOVRYNSLRKAHIFGTFSETOAM 948
QY 481 IFSSV--CYWTLGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIYONPNIVNSV--- 533
DB 949 MYFSYAGCF-----RFG---AYLVAKHLMSEFDVLLVFSAVVFGAMAVQVSSSF 994
QY 534 -----VALSTAGVLVSGFELRNIOEMPIFKIISYFFQKYCSEILVYNEFYGLNFTCG 588
DB 995 APDYAKAKISAAHIIM-----IIKTPL-----IDYSTEGIMPNTELEG--NVTFG 1038
QY 589 SSNSVTNPMCAFTOGI 606
DB 1039 EVVFNYPTRPDPIDPVLQGL 1056

RESULT 5

US-08-784-649A-2
Sequence 2, Application US/08784649A
Patent No. 5830697
GENERAL INFORMATION:
APPLICANT: Sikic, Branislav I
APPLICANT: Chen, Gang
TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
TITLE OF INVENTION: CYCLOSPORIN MODULATION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J

```

; REGISTRATION NUMBER: Reg.No. 5830697 36,677
; REFERENCE/DOCKET NUMBER: 06037/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-322-5070
; TELEFAX: 415-854-0875
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1279 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-784-649A-2

```

```

Query Match 7.2%, Score 240; DB 2; Length 1279;
Best Local Similarity 20.7%, Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

```

QY 41 LHASYSVSHRYRPMWDTSCROMTROIKLDVSLYVSGQIMCTIGSSGSKTLLDAMS 100
DB 396 VHFSPSRKEVK-----ILKGLNLKVGSGQVVALVGNCGCKSTTVQLMQ 440
QY 101 GLRLRAGTFLGEVYVNGRALR--REQFQDCSYVLSQSDTLSSLTVRETLHTYLLAIR 158
DB 441 -RL--YDPTGMSVSDQDITRTINVRFLREITIGVSOEPLVFAATTIAENIRY----- 489
QY 159 GNPSPQKVEAVAAE-----LSLSVADRLIGNTSLSGISGERRRYSIAQILODP 211
DB 490 GRENVTDEIEKAVKEANAYDFIMKLPKKEPTLVGENG--AOLSGGOKRIATARALVNP 548
QY 212 KVMLEDEPTGLDQMTANQIVLVLLVELARNRRIYVLTIHOPRSELPOLFKAIALISFGEI 271
DB 549 KILLDEATSLD--TESAVVQVALDKARKGRITIVIAH--RLSTVRNADVIAGDDCVI 605
QY 272 IFGCTPRMELDFNDGCPCEPHSNPFDYMDLTSVTQSKEREIETS-----KRVQ 323
DB 606 VEKGNHDELM-----KEKGIYFKLVMTQAGNEVELENADESKSEIDALE 651
QY 324 MIESAYKKSATCKHTLKNIERMKHL-----KTLPMVPE-----KTRDSPG-- 363
DB 652 MSSNDSRSSILRKSTRSRVSGSOADRKSLTKREALDESIPVPSFMRIMKLNLTMPYEV 711
QY 364 -----VESK-LGVLLR-----RVTRN-----LVNRKLAVITRL 391
DB 712 VGVECAIINGGLQPAFAIIFSKIGVFTRIDDPETKRQNSMLFSLFLALGIIISFTFFL 771
QY 392 QNLIMG-----LFLLEFVLAVRSNV-----LKGAIODR 419
DB 772 QGFTFGAGELITRLRYMFRSMLRQDVSWFDDPKMTGTGALTTRLANDAAQVGAIGSR 831
QY 420 V-----GLXQVQATPYTGLNANVNLFPVL-----RAVDSQ 454
DB 832 LAVITQMIANLGTGIIISFTIGMQLTLLAI--VPIIATAGVEMKMLSGQALKDKKEL 889
QY 455 DGL-----YOKQMMLAYALAHV-----LPEFVAVATIE 482
DB 890 EGAGKRIATEALEINFRVYSLTQEOKEFEMTAQSLQVRYRNSLRKAHIFGIFESTQAMMY 949
QY 483 SSV--CWITGLHBEVARFGFSAALLAPHLIGEP--LTLVLLGIQONPNIVNSV----- 533
DB 950 FSYAGCF-----RFG--AYLVAKHLMSEFVLLVFSAVVGAAMAAGVYSSRAP 995
QY 534 ---VALLSAGVLVSGFLRNIOEMPIPKIISYTFQKCYSELIVVNEFGALNFTQSS 590
DB 996 DYAAKAKISAAHIIN-----IIEKTPJ-----IDSYSTEGMLPMTLEG--NVTGGEV 1039
QY 591 NWSVTTPMCAFTQGI 606
DB 1040 VFNPTRPDIPVLOGL 1055

```

RESULT 6
US-08-583-276-19

```

; Sequence 19, Application US/08583276
; Patent No. 5837536
; GENERAL INFORMATION:
; APPLICANT: McDonagh, Kevin T.
; APPLICANT: Niemiuis, Arthur
; APPLICANT: Tolstoshev, Paul
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
; TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
; TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DM4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,276
; FILING DATE: 05-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/332,444
; FILING DATE: 31-OCT-1994
; APPLICATION NUMBER: 07/887,712
; FILING DATE: 22-MAY-1992
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; US-08-583-276-19

```

```

Query Match 7.2%, Score 240; DB 2; Length 1280;
Best Local Similarity 20.7%, Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

```

QY 41 LHASYSVSHRYRPMWDTSCROMTROIKLDVSLYVSGQIMCTIGSSGSKTLLDAMS 100
DB 397 VHFSPSRKEVK-----ILKGLNLKVGSGQVVALVGNCGCKSTTVQLMQ 441
QY 101 GLRLRAGTFLGEVYVNGRALR--REQFQDCSYVLSQSDTLSSLTVRETLHTYLLAIR 158
DB 442 -RL--YDPTGMSVSDQDITRTINVRFLREITIGVSOEPLVFAATTIAENIRY----- 490
QY 159 GNPSPQKVEAVAAE-----LSLSVADRLIGNTSLSGISGERRRYSIAQILODP 211
DB 491 GRENVTDEIEKAVKEANAYDFIMKLPKKEPTLVGENG--AOLSGGOKRIATARALVNP 549
QY 212 KVMLEDEPTGLDQMTANQIVLVLLVELARNRRIYVLTIHOPRSELPOLFKAIALISFGEI 271
DB 550 KILLDEATSLD--TESAVVQVALDKARKGRITIVIAH--RLSTVRNADVIAGDDCVI 606
QY 272 IFGCTPRMELDFNDGCPCEPHSNPFDYMDLTSVTQSKEREIETS-----KRVQ 323
DB 607 VEKGNHDELM-----KEKGIYFKLVMTQAGNEVELENADESKSEIDALE 652
QY 324 MIESAYKKSATCKHTLKNIERMKHL-----KTLPMVPE-----KTRDSPG-- 363
DB 653 MSSNDSRSSILRKSTRSRVSGSOADRKSLTKREALDESIPVPSFMRIMKLNLTMPYEV 712
QY 364 -----VESK-LGVLLR-----RVTRN-----LVNRKLAVITRL 391
DB 713 VGVECAIINGGLQPAFAIIFSKIGVFTRIDDPETKRQNSMLFSLFLALGIIISFTFFL 772

```

OY 392 QNLIMG-----LFLFFVLRVRSNV-----LKGAIODR 419
DB 773 QGFTFGAGELTRKRLRYMFRSMLRODVSWFDDPKNTGTALTRLANDAAQVGAIGSR 832
OY 420 V-----GLYQVGAATPYTGMLNANVLPVL-----RAVSDQSO 454
DB 833 LAVTONTIANLGTGIIISFYGMQLTLLAI--VPITAIAGVEMKMLSGQALKDKKEL 890
OY 455 DGL-----YQWQMLAYALHV-----LPFSVATMIF 482
DB 891 EGAGKITEAIEENFRVVSLOBOKEFHMTAQSLOVYRNSLRKAHIFGITFSFTQAMMY 950
OY 483 SSV--CYWTGLHPEVARFGYFSALLAPHLIGEF--LTVLGLIYQNPINVSV-----533
DB 951 FSYAGCF-----RFG-----AYLVAKHLSFEDVLLVFSAVFGAMAVQVSSFP 996
OY 534 ---VALLSTAGVLVSGFLRNIOEMPPEKTIISFTFOKYCSEILVNEFYGLNFTCGSS 590
DB 997 DYAKAKISAHHIM-----IIEKTP-----IDYSTEGIMPTLEG--NATFGEV 1040
OY 591 NNSVTTPMCAFTQGI 606
DB 1041 VFNYPTRPDIPVLQGL 1056

RESULT 7
5206352-4
Patent No. 5206352
Applicant: Ronlison, Igor B.; Pastan Ira H.; Gottesman,
Michael M.
TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/622,836
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 892,575
FILING DATE: 01-AUG-1986
APPLICATION NUMBER: 845,610
FILING DATE: 28-MAR-1986
SEQ ID NO: 4:
LENGTH: 1280
5206352-4

Query Match 7.2%; Score 240; DB 6; Length 1280;
Best Local Similarity 20.7%; Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

OY 41 LRAYSVSHRVPRMMDITSCROQWTRQILKDVSLVESGOIMCIIGSSSGKTLLDAMS 100
DB 397 VHFSPSRKEV-----ILKGLNLVQSQVVALVNSCGKSTTVQLMQ 441
OY 101 GRLGAGTFLGEEVYNGRAL--REOFODCFSYVLOSPTLSSLVRETILHTALLAIR 158
DB 442 -RL--YDEPTEGAVSDGODIRTNVRLREIIGVVSQEVLEFATIAENIR-----490
OY 159 GNPFSQKKEVNAE-----LSLSHVAURLJNYSIGISTGERRRVSTIAQLODP 211
DB 491 GRENVTMOIEKAVEANAYDFIMKLPKRFDTLVGERG--AQLSGQOKORIALARLVNP 549
OY 212 KYMLFDEPTGLDCQTANOIVVLVLELARNRIVLTIHOPSELFQFDKALISFBL 271
DB 550 KIILDEATSLD--TESAVVQVADKARKGTTTIVAH--RLSTVRNADYIAGDDYI 606
OY 272 IFCGTPAEMLDFNDGCPCEPHSNPFEDYDLTSVDTSQEREIETS-----KRVQ 323
DB 607 VEKGNHDELM-----KEKGIYFKLVLMOTAGNVELENNADESKSIDALE 652
OY 324 MIESAYKSAICHKTIKRIEMKHL-----KTLPMVVP-----KTKDSPG--363
DB 653 MSSNDRSLLTRKSTRSVSGQAODRKLSTKEALDESIPVSWFRIMKINLTFEMPYEV 712

OY 364 -----VESK-LGVLLR-----RVTRN-----LVNRKLAIVTRLL 391
DB 713 VGVECAIINGLOPAFAIIFSKIGVTRIDDPETKRQNSNLSFLFLALGIIISFITFEL 772
OY 392 QNLIMG-----LFLFFVLRVRSNV-----LKGAIODR 419
DB 773 QGFTFGAGELTRKRLRYMFRSMLRODVSWFDDPKNTGTALTRLANDAAQVGAIGSR 832
OY 420 V-----GLYQVGAATPYTGMLNANVLPVL-----RAVSDQSO 454
DB 833 LAVTONTIANLGTGIIISFYGMQLTLLAI--VPITAIAGVEMKMLSGQALKDKKEL 890
OY 455 DGL-----YQWQMLAYALHV-----LPFSVATMIF 482
DB 891 EGAGKITEAIEENFRVVSLOBOKEFHMTAQSLOVYRNSLRKAHIFGITFSFTQAMMY 950
OY 483 SSV--CYWTGLHPEVARFGYFSALLAPHLIGEF--LTVLGLIYQNPINVSV-----533
DB 951 FSYAGCF-----RFG-----AYLVAKHLSFEDVLLVFSAVFGAMAVQVSSFP 996
OY 534 ---VALLSTAGVLVSGFLRNIOEMPPEKTIISFTFOKYCSEILVNEFYGLNFTCGSS 590
DB 997 DYAKAKISAHHIM-----IIEKTP-----IDYSTEGIMPTLEG--NATFGEV 1040
OY 591 NNSVTTPMCAFTQGI 606
DB 1041 VFNYPTRPDIPVLQGL 1056

RESULT 8
US-08-665-259-25
Sequence 25, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-3415
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1684 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

Query Match 7.2%; Score 239.5; DB 3; Length 1704;
Best Local Similarity 32.5%; Pred. No. 7.5e-16;
Matches 76; Conservative 43; Mismatches 100; Indels 19; Gaps 9;

QY 66 ROLKDVSLVSGQIMCITLSSGSGKTTLLDMSGRGAGTFLGEVYNGRALRR-- 123
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 546 RAAVRDNLNLYEQIVLGHNGAGTTLTSMITGEPPTS--GRAVYSGYEISODMV 602
QY 124 QFODCFEYVLOSOTLLSLVRETLHTALAIRGNGPSFOKKVEAVMAELISHVADR 183
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 603 QIRKSLCICPOHDLFNLVVAELHYAQL---KGL--SRQCPREYKQMLHITGLEDK 657
QY 184 LIGNYSLIGISTEERRRVSIAAQLLDDPKVLFDEPTTGIDCMTANOIVLVLELARNR 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 658 --WNSRSRFLSGMRRRLSIGIALIGSKVYLILDEPTSGDAISRRAIMDL-QROKSDR 714
QY 244 IVVLTTHQPSSELFQDKTALISFGLIFCGTPAEMLDFFNDCGYP---PEHSNPF 299
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 715 TIVLTTHF-WDEADLDGRITAMAKGELCCGSSLELKORYG-AGYHMTLVKPEHCNPF 772

RESULT 11

US-09-134-001C-3832
; Sequence 3832, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3832
; LENGTH: 242
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

Query Match 6.9%; Score 230.5; DB 4; Length 242;
Best Local Similarity 26.1%; Pred. No. 2.3e-16;
Matches 64; Conservative 56; Mismatches 90; Indels 35; Gaps 7;

QY 67 QILKDVSLVSGQIMCITLSSGSGKTTLLDMSGRGAGTFLGEVYNGRAL----- 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 17 EYLRDNLNLYEQIVLGHNGAGTTLTSMITGEPPTS--GRAVYSGYEISODMV 602
QY 121 ---RREQDFCFEYVLOSOTLLSLVRETLHTALAIRGNGPSFOKKVEAVMAEL 175
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 68 TQHNHDLNLRKMGKMGVQFQNLFPKHYIENVMALPL-LKDSKQDLKERALYLEEV 126
QY 176 SLSHVADRILIGNYSLGISTEERRRVSIAAQLLDDPKVLFDEPTTGIDCMTANOIVVL 235
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 127 GLKDKAD---SYR-NQSGGKQVAVARALAMEPDMVLFPEPFSALDPVAVGDLKVM 181
QY 236 VELARRNRIVLTTHQPSSELFQDKTALISFGLIFCGTPAEMLDFFNDCGYPCEHS 295
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 182 ROLANESMTWIVTVE-MNFAKEISDKVFMADGVVESGTPONT-----FENPOHS 232
QY 296 NPFDF 300
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 233 RTENF 237

RESULT 12
US-09-134-001C-3369
; Sequence 3369, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3369
; LENGTH: 460
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3369

Query Match 6.9%; Score 229; DB 4; Length 460;
Best Local Similarity 25.2%; Pred. No. 1e-15; Mismatches 121; Indels 42; Gaps 10;

QY 66 ROLKDVSLVSGQIMCITLSSGSGKTTLLDMSGRGAGTFLGEVYNGRALRR-- 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 57 KRAVDNISLDIOGSEFIATIGTSGSKTAL-RMIRMLENTD-QQIMANGKDVANNMP 113
QY 123 EPODFCFEYVLOSOTLLSLVRETLHTALAIRGNGPSFOKKVEAVMAELISHVADR 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 114 VELRRSIGVYDIOIGLMPHTIREN---IVLVPLTLKWSREKDEKAKELIKVLDPE 168
QY 183 RLIGNYSLIGISTEERRRVSIAAQLLDDPKVLFDEPTTGIDCMTANOIVLVLELARNR 242
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 169 EYLRDNP-ALSGGQQRIGVYRALAEODITLMBEPGALPITVDTQDLVKEIQKL 227
QY 243 RIVVLTTHQPSSELFQDKTALISFGLIFCGTPAEMLDFFNDCGYPCEHSNPFDFYM 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 228 GTFIEFVTHDMDEATIKADKICIMSKGVQYDTPDNL-----RYRPN-----DFVR 275
QY 303 DLTSVDQSKEREIEISKVQVMEISAKSAICHTLK-----NIERKKHLKTLPM 353
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 276 DFT---GQNLIDRPNMKSVESAMIKPV---TYKADSLDVAVIMRRRVDTIV 326
QY 354 VPEKTR 359
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
DB 327 VANNOK 332

RESULT 13
US-08-395-246C-2
; Sequence 2, Application US/08395246C
; Patent No. 5773214

; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,246C
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas G.
 REGISTRATION NUMBER: 35784
 REFERENCE/DOCKET NUMBER: X9683
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-2459
 TELEFAX: 317-277-1917
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1307 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-395-246C-2

Query Match 6.9%; Score 229; DB 1; Length 1307;
 Best Local Similarity 23.0%; Pred. No. 6,4e-15;
 Matches 127; Conservative 110; Mismatches 188; Indels 128; Gaps 29;

```

OY 68 ILKDVSLVYESGOIMCIGSSGSGKTTLLDMSGRIGR-AGTFLGEVY-----VNGRALRR 122
DB 407 VAHDLSCYIPACKTAFAVGPSSGSKSTISLERFYDPAVAGTMDGHDIGTLRLRLAQ 466
OY 123 EGFQCFSVLSDTLSSLYRETLHTALAIRNGNSFO--KKEAV--MAE----- 174
DB 467 Q-----MSLVSEPRLEFAT-TIAENIRY-GIIGSFREKESYEIRKREVAARMANAHDF 519
OY 175 -LSLSHVADRLIGNTSLGISTGERRRVSIQAQLDPPKMLFDEPTGIDCMTANOIV 233
DB 520 IVALNGDYTNIESLSLGS--GOKORAIARAIITDKPILDLDETSALD--TKSEKLVQ 575
OY 234 LVELARRNRIVLTTHQPRSELFDLFDKIALISFGLFCSTPAEMLDFFNDGCPCE 293
DB 576 ALDKASKGRITIVIAH--RLSTIQKAVNIIVLANGQIEGPHHEHMD----- 622
OY 294 HSNPDFYMDLTVTOSKERELFETSKRVOMIESAYKSAICHKLKNIERNKHLKTLPM 353
DB 623 ---RRCITCDMVEAH-EIKKRYSRSKRSOL-----LTNLSP-----KHPNM 661
OY 354 VPFKTKDSPG-----VF-----SKLGVLRRVTRNLYRNKLAVITFLQLNLGLFL 401
DB 662 TFFPFKQVPGDESDIYISLSDADIGLHGEKRPVSRSLSLIMOPVEEAVSFMTL 721
OY 402 -----FVLNRVSNVLKGAIDRVAGLLYQFVGATPYTGMLNANVLFPVLRV 448
DB 722 FKFLASFNRPEWPFLLGLCASILAGIOPSOAVLFAKAVSTLSLPLE---YKRLR-- 775
OY 449 SDQESDGLYQKQWMLVALHVLRFVSVAIMI-FSS--VCY-----WTGLHPEVARF 499
DB 776 -----HDANFMCMLFMIGIVSLVLYSVQGLTFAISSEKMYRRASQAFLVILHODISFF 830
OY 500 -----GYFSAALIA--PHLIGEFLLVILG--IVONPNIVNSV-----VALLSI 539
DB 831 DQDENTGALTATLSAGTKRELTG--ISGVTLGTLIIYSVNLVASLGAVALVIGMKLALVCI 888
OY 540 AGV--LVGSGFLR 550
DB 889 SAVPALLMCGFVR 901

```

RESULT 14
 US-09-134-001C-3641
 Sequence 3641, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134, 001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3641
 LENGTH: 273
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3641

Query Match 6.6%; Score 220.5; DB 4; Length 273;
 Best Local Similarity 24.7%; Pred. No. 3.5e-15;
 Matches 62; Conservative 52; Mismatches 82; Indels 55; Gaps 7;

```

OY 66 ROIKDVSLVYESGOIMCIGSSGSGKTTLLDMSGRIGRAGTFLGEVYVNGRALRRDF 125
DB 31 KAVLEININKINQGEFLAIVGPGAGKSTLLKVI--LELPIQKGEIIVDCKPFGKNS 87
OY 126 QDCFSVLD-----SDTLSSLYRETLHTALAIRNGNSFOKKEAVNM 172
DB 88 SLKISIVSQKASAFNAPGASVKEVLSGLTITKL-----FQFNKNDYQ-KYTKVL 139
OY 173 AELSLSHVADRLIGNTSLGISTGERRRVSIQAQLDPPKMLFDEPTGIDCMTANOIV 232
DB 140 ERLNISHLT-----NKNIAELISGGQOQRYLIRALISPSVVLDEPTNGIDAKHVSFY 194
OY 233 VLLVELARRNRIVLTTH-----QPRSELPFLDKIALISFGL 271
DB 195 DFLDKLKKOGITIIIVTHTDIGNVADTAEVACLNRHLPHGSTEAFKSLIDEVEISK----- 250
OY 272 IFGCTPAEMLD 282
DB 251 -IYGHPIQFVD 260

```

RESULT 15
 US-08-665-259-27
 Sequence 27, Application US/08665259
 Patent No. 6028173
 GENERAL INFORMATION:
 APPLICANT: Landes, Gregory M.
 APPLICANT: Burn, Timothy C.
 APPLICANT: Connors, Timothy D.
 APPLICANT: Dackowski, William R.
 APPLICANT: Van Raay, Terence J.
 TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
 TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
 NUMBER OF SEQUENCES: 73
 CORRESPONDENCE ADDRESS:
 ADDRESS: GENZYME CORPORATION
 STREET: One Mountain Road
 CITY: Framingham
 STATE: Massachusetts
 COUNTRY: United States of America
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/665,259
 FILING DATE: 17-JUN-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Dugan, Deborah A.
 REGISTRATION NUMBER: 37,315
 REFERENCE/DOCKET NUMBER: 1G5-9.1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 27:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1457 amino acids

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:09:57 ; Search time 22 seconds
(without alignments)

2844.705 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326 1 MGDLSLTPGSGMGLOVNRG.....PALVILGIYFKIRHDLISR 651

Sequence: BL050M62
Gapop 10.0 , Gapext 0.5

Scoring table: 283224 segs, 96134422 residues

Searched: 283224 hits satisfying chosen parameters: 283224

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	649.5	19.5	725	2	ABC transporter-11
2	627	18.9	1294	2	ATP-dependent perme
3	621	18.7	1049	1	ABC transporter-11
4	614	18.5	720	2	probable ABC trans
5	610.5	18.4	725	2	ABC transporter-11
6	608	18.3	708	2	white protein - fr
7	602.5	18.1	687	1	protein F12M16.17
8	597	17.9	590	2	hypothetical prote
9	595.5	17.9	658	2	probable ABC trans
10	591.5	17.8	740	1	hypothetical prote
11	590.5	17.8	740	1	probable ATP-bind
12	589.5	17.7	638	2	white homolog - hu
13	584	17.6	755	2	probable ABC trans
14	583.5	17.5	687	2	hypothetical prote
15	578.5	17.4	646	2	ATP binding cass
16	575.5	17.3	649	2	probable ABC trans
17	572	17.2	739	2	ATP transporter-11
18	571.5	17.2	609	2	probable ABC trans
19	564	17.0	662	2	ABC transporter-11
20	554	16.7	559	2	probable ABC trans
21	552.5	16.6	635	2	ABC transporter-11
22	545.5	16.4	608	2	protein COSD10.3
23	540	16.2	678	2	hypothetical prote
24	539	16.2	577	2	hypothetical prote
25	529.5	15.9	653	2	hypothetical prote
26	527.5	15.9	659	2	hypothetical prote
27	517.5	15.6	610	2	hypothetical prote
28	517.5	15.6	630	2	hypothetical prote
29	514.5	15.5	705	2	protein C10C6.5
					probable ABC trans

30	504	15.2	695	2	T21109	hypothetical prote
31	497	14.9	547	2	T31543	hypothetical prote
32	488	14.7	737	2	T46101	ABC transporter-11
33	464	14.0	1450	2	A84780	probable ABC trans
34	456.5	13.7	1426	2	T30567	ATP-binding cass
35	454.5	13.7	675	1	FYFPR	ATP-binding cass
36	435	13.1	668	2	S55023	brown protein - fr
37	429	12.9	1469	2	H96622	brown protein - fr
38	425.5	12.8	1443	2	T02491	probable ABC trans
39	425	12.8	1451	2	B86286	probable ABC trans
40	421	12.7	1450	2	T45888	F9L1.15 protein -
41	420	12.6	1435	2	D96693	ABC transporter-11
42	418.5	12.6	1435	2	T02644	protein putative A
43	417.5	12.6	1619	2	T30541	ABC-type transport
44	413	12.4	1413	2	G84790	ABC1 transport pro
45	410	12.3	1177	1	D71416	probable ABC trans
						probable PD5-1ike

ALIGNMENTS

RESULT 1	
ABC transporter-like protein - Arabidopsis thaliana	
N: Alternate names: protein T26112.10	
C: Species: Arabidopsis thaliana (mouse-ear cress)	
C: Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000	
C: Accession: T47652	
R: Montfort, A.; Casacuberta, E.; Puigdomenech, P.; Meyers, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, February 2000	
A: Reference number: Z24471	
A: Accession: T47652	
A: Status: preliminary	
A: Molecule type: DNA	
A: Residues: 1-725 <MON>	
A: Cross-references: EMBL:AL132954	
A: Experimental source: cultivar Columbia; BAC clone T26112	
C: Genetics:	
A: Map position: 3	
A: Note: T26112.10	
C: Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12M16.1;	
Query Match	19.5%; Score 649.5; DB 2; Length 725;
Best Local Similarity	29.4%; Pred. No. 1e-40;
Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;	
QY 33 PEPHSGILHASYSHRYRPMWMDITSCROQTROIADVSLYVSGQIMCLGSSGSK 92	
DB 68 PVPYVLRNENLDYDTLRR-----FGSRQNGVKTLDDVSGEASDGLAVLGASGAK 123	
QY 93 TLLDAMSGRLGACFTLGEVYVNG-RALRREPOFDCFSYVQSPILSLIVRETLYT 151	
DB 124 STLIDLAGRYAE-GSLRGSYVLNGBKYLQSLKLVISAYVQDDLPMLVKEITMRA 182	
QY 152 ALLAIRRG-NRGSFQKRYEAYAEISLHVADRLIGNSLGISIGERRRVSIAAQLD 210	
DB 183 SEFRPLRSLSKKRREREALELDGLRMAATVGDGHRGVSGERRRVSIGIDIHD 242	
QY 211 PKVMLDEPTTGIDCMATNOIVLVLLVELARNRIVLVTHOPRSELPQFQKIALISPE 270	
DB 243 PIVLELDPTSGLDSTNAPMVVQVLRKLAQSGSYIMSIHOPSAIVELDLRLILSRK 302	
QY 271 LIFCGTPAEMLDFFNDGCPCEPESHNPDEFYMDLTVDTQSKREIETS----- 319	
DB 303 SVFNQSPASLPQGFDFSRPIPEKENSIEFALDY-----RELQSGNEKRALVDPN 354	
QY 320 -----KRYQIESAIV-----KSAICHTL--KNIERKHLKTLPMVPFKTKD 360	
DB 355 EKMOONKISLQSAVQTKLDDQRLSLKAEAINASVSRGKLVSGSSRSNPTSMETVSSA 414	
QY 361 SPGVESKGLVLRVTRNLVKNKLAVITRLLQNLMLGFLFFVLRVSNVYKGLQIDRV 420	
DB 415 NPSLFEPE-ILAKRIKMKWIRPPELVGTRIAVWVTC-LLATVYWKLDHPTPRGA-QERTL 471	

QY 421 GLTYQVQATPTTGMLNANVLEPVLRANVSDSDGLYOKOMMLAYLHVLPSVSATM 480
 Db 472 -TLFAVYVPMETCCIDNVPVFIQERYIFLRETHNARVSSVYSHSLVSLPOLAPSL 530
 QY 481 IFSSVCYWTGILHPEVARGYSALAPHLIGELTVLIGVQNPVMS-VYALLSI 539
 Db 531 VFSATTEVTGLSGGSEGFVYCLITVASFWSGSSVTFISGV--PIMICMYSTIYL 588
 QY 540 AGVLVSGFLNIOEPIPKIISYTFQKCYCELIIVNVEYGLNFTGSSNSVSTPM 599
 Db 589 AYCLLSGFYVNRDIRPPTWTFHISLIKTPYEAVLINEF-----DDPS 633
 QY 600 CAFTOGIOFIETKCPGATS 618
 Db 634 RCFVGVQVFDSTLLGVVS 652

RESULT 2

Probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical
 C:Species: Saccharomyces cerevisiae
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
 C:Accession: S77690; S66767; S66768
 R:Alexandrak, D.; Katsoulou, C.; Tzermla, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66756
 A:Accession: S77690
 A:Molecule type: DNA
 A:Residues: 1-1294 <ALE>
 A:Cross-references: EMBL:Z74816; MIPS:YOL075c
 A:Note: this is a revision to the sequence from reference S66756
 A:Accession: S66767
 A:Molecule type: DNA
 A:Residues: 1-179 'TTRTGVLYVKRED' <ALM>
 A:Cross-references: EMBL:Z74816
 A:Experimental source: Strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Note: this was assumed to be protein YOL074c
 A:Accession: S66768
 A:Molecule type: DNA
 A:Residues: 200-1294 <ALF>
 A:Cross-references: EMBL:Z74817
 A:Experimental source: Strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Note: this was assumed to be the complete sequence of protein YOL075c
 C:Genetics:
 A:Cross-references: SGD:S0005435
 A:Map position: 15L
 A:Note: YOL075c
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:376-382/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:606-622/Domain: transmembrane #status predicted <TM4>
 F:710-916/Domain: ATP-binding cassette homology <ABC2>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1042-1058/Domain: transmembrane #status predicted <TM5>
 F:1125-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.98; Score 627; DB 2; Length 1294;
 Best Local Similarity 31.7%; Pred. NO. 1,1e-38;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

QY 65 TROLKDVSLYVSGOIMCIGSSGKTLIDAMGRGRATFL-----GEVYVNGRA 119
 Db 706 TKELIOSVNAIFKPGMINIMGPSGSKSLNLWISGR-LKSSVFAKFDPTSGSIMFNID 764

QY 120 LRROFDCFSYVLQSDT-LLSLTFRETLHYTALLARGNPGSFQKVEANAEISSL 178
 Db 765 VSELMFKVNCYVSSODDHLAALVTKTLTYAALRHHLTEERMERTNLRSGLK 824
 QY 179 HVADRLIGNYSGLISTGERRRVSIAADLPYKMLDEPPTGLDCMTANQIVLLVEL 238
 Db 825 HCENNIIGNERVKGISGGERKRVTMGVQLNDPILLLDEPTSGLDSTSAITILEKL 884
 QY 239 AR-RNRIVVLTIRHPRSLFQFLKIALIS-FGLIFCGTPAEMLDFNDCGCPERSN 296
 Db 885 CREGKTLITIRHPRSELFRPGVNLIIAASGRFAENGSPDEMAVTELEGNCPSFTN 944
 QY 297 PFDFYMLDTSVDOSKEREIETSKRVOMIESAYKSAICHITIKIEMKHLKTPMPF 356
 Db 945 VADFFLLISVNTONEQNEISSRAVEKILSAMK-----AMDN-ESLSPTISEK 994
 QY 357 KTKDSPGVSKIGVLRREVTNLY-----RNKLAVITLLNLNIGLFL 401
 Db 995 QOYQSESEFTLEYSEVRK-PANVLVAVLVNVRQFTTTRRSPDSLMARIADIPGLGVFA 1053
 QY 402 FFVLRVNSNVLKGAIDRRVGLLYQVCGATPYTGMLNANVLEPVLRANVSDSDGLYQK 461
 Db 1054 LFFAPVKNHYT--SISNRLGIAQEST-ALYFVGMGLNACYPTERDYEEYNVYVIA 1110
 QY 462 QMNLAYLHVPESVVAITMISVYWTGLGHPVARGYSAALLAPHLI---GEFLTL 518
 Db 1111 PFLAVYMTLELPISALSVLAFTVLAAGI-PRTA--GNFPAVYCSFIYTCGERLGI 1167
 QY 519 VLLGIQVNPV-IVNSVALLSIAGVLVSGFLRNIOEPIPKIISYTFQKCYCELIIV 577
 Db 1168 MTNFFRPPRGVAVNCISIIISIGOMGSLMST---GSRVLRKGFNTLNPVGYTSMIITN 1223
 QY 578 NEFTG-LNFTC--GSSNVSVTTPMCAFTOG 605
 Db 1224 FAFPGNLKTCEDCGKNSDGT---CEFANG 1250

RESULT 3

ATP-dependent permease ADP1 precursor - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YCR011c; protein YCR105
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 C:Accession: S19421; S40914
 R:Goffeau, A.; Purnelle, B.; Skala, J.
 submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19420
 A:Accession: S19421
 A:Molecule type: DNA
 A:Residues: 1-1049 <GOF>
 A:Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42328.1; PID:g1907154; GSPDB:
 R:Purnelle, B.; Skala, J.; Goffeau, A.
 Yeast 7, 867-872, 1991
 A:Title: The product of the YCR105 gene located on the chromosome III from Saccharom
 A:Reference number: S40914; MUID:92160395; PMID:11789009
 A:Accession: S40914
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1049 <PUR>
 R:Skala, J.; Purnelle, B.; Goffeau, A.
 Yeast 8, 409-417, 1992
 A:Title: The complete sequence of a 10.8 kb segment distal of SUR2 on the right arm
 K genes.
 A:Reference number: S25353; MUID:92327849; PMID:1626432
 A:Contents: annotation
 A:Gene: SGD:ADP1; MIPS:YCR011c
 C:Genetics:
 A:Gene: SGD:ADP1; MIPS:YCR011c
 A:Cross-references: SGD:S0000604; MIPS:YCR011c
 A:Map position: 3R
 C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>

Query Match	18.7%;	Score 621;	DB 1;	Length 1049;
Best Local Similarity	28.6%;	Pred. NO. 2.3e-38;		
Matches 196;	Conservative 111;	Mismatches 223;	Indels 156;	Gaps 22

Db 1015 -VVQNLVFDIKILALFNVLIMGYL 1039

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

Query Match	18.5%	Score 614;	DB 3;	length 720;
Best Local Similarity	28.7%	Pred NO. 4.6e-38;		
Matches - 182;	Conservative 122;	Mismatches 254;	Indels 76;	Gaps 17;

Db 623 EF-----SDPTECFVRGVQLFDNS 641

R; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.
M.; Koo, H.; Moffat, K.S.; Cronin, T.A.; Shear, W.; Varley, G.P.; ...

Db 646 GLVALKIR 653

355 СТРБИ ТЕНИКВОНУОСНУОДНУ ЕДНОСЛОВНУ МЕРНА ТАЈ СТОПОВИЦА СЛОЖЕНА

```
QY 347 ----HLKTLPMVPEKTKDSQVFSKGLVLRRTVNLNKLAVITRLQNLINGLELLE 402
DB 415 GATTTTTLTAVAFANP-----MMIEIKLSKRSMLNSRQPELFGIRIASVITG-FILA 469
QY 403 FLVLRASNVLKGAIDRGLQYFGVATPYTGMNANVLPFLVLRASVQESODGLQYKQK 462
DB 470 TVFRLDNPCKG-VQERLG-FEAFAMSTWETTCALAPVLOERYIFMNETVYNAVRSS 527
QY 463 MMLAVALHVLPEPSVAVATMIFSSVCYWTGLHPEVAFGFSALLAPHLIGELFLVILG 522
DB 528 YLSHAIVSFPSELIFLSAFAATYTNVAGLOGGLGLFLYCLIIILASWSSGFVTFISG 587
QY 523 IVQNNIV---NSVALLSIAGVLVSGFPLNIQEMPIPKIISYTFQKCYCEILVYNE 579
DB 568 VV-PSVMLGYTIVAIL-AVFLFSGFEFINRNRIPDYIMFHMVMSLVKPYEAVLQNE 643
QY 580 FYGLN-FTGC-----SSNVSTTNMCAFTQG----- 605
DB 644 FSDATKCFVKGVOIFDNTPLSELPEVMKLLGLTVSKSLGYTISSTGLTGSDDLROG 703
QY 606 -IOPLEKTCGATSRFTNPLILYSFIPALVILG 638
DB 704 VVOLSKMNCLEITVAFGFFFLIFYF---TLLIG 734
```

RESULT 11

```
G02068
white homolog - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001
C:Accession: G02068
R:Clomp, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Archiniegas,
submitted to the EMBL Data Library, August 1995
A:Reference number: H00769
A:Accession: G02068
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <CRO>
A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277
C:Genetics:
A:Gene: white
C:Superfamily: fruit fly white protein; ATP-binding cassette homolog
C:Keywords: ATP; nucleotide binding; P-loop
F:61-253/Domain: ATP-binding cassette homology <ABC>
F:78-85/Region: nucleotide-binding motif A (P-loop)
```

Query Match 17.8%; Score 590.5; DB 2; Length 638;
Best Local Similarity 26.3%; Pred. No. 2.3e-36;
Matches 164; Conservative 142; Mismatches 266; Indels 51; Gaps 14;

```
QY 44 SYSVSHRRPMDITSCROOQWROIILKDVLYVESGOIMCIGSSGSKTLLDAMGRL 103
DB 43 SYSVPE--GPRW-----RKGKYLTLGLISGKENSGLVAMGSGAKSTLMIILNGY- 94
QY 104 GRAGTFLGEVYVNGRALRRQFQDCFSYVLOSDTLLSLVRETILHTALAIIRGNPGS 163
DB 95 -RETGMGAVLNLPRDLRCFRKVSXYIMODMLLPHLVYQEMAMVSAHLKQDEKDEGR 153
QY 164 FQKVEVAMELISHVADRILGNYSLGIGSTGERRRVSIAPQLODPKVMLEDEPTTGL 223
DB 154 -RENVKELITLALGLSCA-----NTRTGSLSGGQKRKLALALELVNPNVAFEPDPSGL 207
QY 224 DCAHTANIIVLVELARNRIVITIHQPSSELPQLQDLKTAIISFGLICGTPAEMIDF 283
DB 208 DSAACFOVYSIMGLAGGSRISICTTHQPSAKLEFLDOLYVLSQGCYVRGKVCNLPY 267
QY 284 FNDGVPCEPSPHNPFDYMDLTSYDQSKREIETSKRYOMIESAYKSA-----I 334
DB 268 LRDGLNCPYTHNPADPVYMEVASGEYDQNSRLVRAVREGMCSDBKRDGLGDAEVNPL 327
QY 335 CHHTLKIERMKHLKTLPMVPEKTKDSQV-----FSKGLVLRRTVNLNKL 384
```

```
DB 328 MHRPEEVAKQTRKLGK-----RKDSSMECHSFSASCLQPCILFRKFLSMDRSV 381
QY 385 AVITRLQNLINGLELLEFVLRVRSNVLKGAIQDRVGLQYFGVATPYTGMNANVLPFV 444
DB 382 LTHLSTHIGIGLGLILGLYLGIGNEAK--VLSNSGLFESMLPIMAPALPVLTPPL 439
QY 445 IRAVSDQESODGLYQKQOMMLAYALHVLPEPSVAVATMIFSSVCYWTGLHPEVAFGFS 504
DB 440 EMGVFLREHLNHWISLAKAYTLAKTMADVPFOIMEFVAICSTIYMTTSSQPSDAVAF 499
QY 505 ALLAPHLIGELFLVILGIVQNPVINSVALLSIAGVLVSGFPLNIQEMPIPKIISY 564
DB 500 LGTWSLVAQSLGL-LIGAASLQVAFVGVATIPVLLPSGFVSDITPTIYQMSY 558
QY 565 FTFOKCYCEILVYNEFGYGLNFTCGSSNVSTTNMCAFTQGIQTEKTCGATSRFTNMF 624
DB 559 ISVYRGEVYLS-ITGLD---REDLHCDIDETCHF-QKSEALIRELDVENAKYLIDF 612
QY 625 ILVSPFIPALVILGIV--FKIR 645
DB 613 IVLGIFISLRLIAFVLRIKIR 635
```

RESULT 12

```
C86441
probable ABC transporter (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86441
R:Thellogis, A.; Becker, J.R.; Palm, C.J.; Federjpiel, N.A.; Kaul, S.; White, O.; A.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewai
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Ki
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Matz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tai
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-646 <STO>
A:Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GNO0141
C:Genetics:
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;
```

Query Match 17.7%; Score 589.5; DB 2; Length 646;
Best Local Similarity 29.9%; Pred. No. 2.8e-36;
Matches 183; Conservative 112; Mismatches 246; Indels 71; Gaps 22;

```
QY 10 GGSN-GLQVNRGSS-SLEGAPATAPSPISGLTHASYS-----HVRPMDITSCR 61
DB 14 GGVAVQGLPDMSPDQSSVLAFFPTTSQ--GLQMSYPTTLKEVYKVK-TEQTSQCM 69
QY 62 QQW---TROIILKDVSLVYESGOIMCIGSSGSKTLLDAMSGRLGAGTFLGEVYNGR 118
DB 70 GSWKSKRKTLLNGITGAVGCEFLAMGPGSGSKTLLSLALGSLK--TFSGVMYNG- 126
QY 119 ALRREQQDF-----SVILOSDTLLSLVRETILHTALAIIRGNPGSFQK-----KVE 169
DB 127 ----QPSGCIKRTGTGAADDVLYPHLIVMETLFEFALLRL-----PSSLRDBKAEHVD 178
QY 170 AVMAELSLSHVADRILGNSLGSISGERRRVSIAPQLODPKVMLEDEPTTGLDCTAN 229
DB 179 RVIAELGLNCTVSMIGPLRFGISGGEKRVISIGEMLINPILLDEPTSGIDSTYAH 238
QY 230 QIVLVELARRNRIVITIHQPSSELPQLQDLKTAIISFGLICGTPAEMIDFNDY 289
DB 239 RIYTTIKRLASGGRTVTTIHQPSSRRIYHFDKAVLLSESPITYYGAASSAVEFFSLGF 298
```

OY 290 PCPEHNPDPFYMDLTS---VDTOSKEKEITSKRVOMIESAKY-----KSAICHTL 339
 DB 299 STSLTNPADLLDLNIPDPDTPKESDOKTVKELTVSAVEKNISTKLKAEICNAES 358
 OY 340 KNIEMK-HKLTLPVPFKTDSPGVSKLGLVLRVYRNL---VRNLAVITRLQMLI 395
 DB 359 HSYETKRAANKLSEQCHT---WVQFVLQKRGYRERRESFNLRLEQVISAFL 414
 OY 396 MGLFLFEVLVRNSVLKGAIDRVGLLYOFVGAIPYGMILNAVLPFLRAVSDQSD 455
 DB 415 GG--LLMW-----HTPKSHIQDRFALLFFSVMGFPLNVAFTFQEKRLIKERS 466
 OY 456 GLYOKOMMLAYALHVPFSVVAWTFSSVCYWTGLGHPVARFGYSALLAPHLIGEF 515
 DB 467 GMYRLSSYFARNGDLPLELALPTAFVFIYMMGGLKPDPTFELSLVLYSVLVAOG 526
 OY 516 LTVLGLVONPNTVNSVVALSTAGVYSGFLRNIOEMPIPKI-ISTYTPKQYSEI 574
 DB 527 LGLFGLALINIKATATLASTVTLVFLIAGTY---VQOLP-PRIVMLKYSYSTYCKL 582
 OY 575 LVVNEFYGLNFT 586
 DB 583 LL-----GLOYT 589

RESULT 13

Probable ABC transporter [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence, revision 02-Feb-2001 #text, change 16-Feb-2001
 C:Accession: G84791
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Unayam, L.; Tallon, L.;
 Nuss, D.; Mierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84791
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-755 <STO>
 A:Cross-references: GB:AE002093; NID:g4056489; PIDN:AAC98055.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: Atg2j37360
 A:Map position: 2
 C:Superfamily: Arabidops thaliana probable ATP-binding cassette protein F12L6.1; ATP-1

Query Match 17.68; Score 584; DB 2; Length 755;
 Best Local Similarity 27.28; Pred. No. 9e-36;

Matches 173; Conservative 128; Mismatches 253; Indels 82; Gaps 19;

OY 21 SOSISLEGAPAT--APEPHSIGILHASYSGHRVPMMDITSCROQM-----TROILKDV 72
 DB 79 SFNSMASAPASSTISSSPVLSFTDITVSVKIQ-KKFPPLACCRSGNDSSVNTKILNGI 137
 OY 73 SLVESGOICILSSSGGKTTLLDMSGRIGRAGTFLGAYVNGRALRREOFODCSYV 132
 DB 138 SGEAREGEMAAVIGASGSKSTLIDALANRIAR-DSLRSITLNGEVLESMSKVISAVY 196
 OY 133 LQSDTLISSTVRETIHYTALLAIRGNPSFOKK-----VEAVMAELSLSHVADRILGN 187
 DB 197 MODDLFPMLTVEETLWFSAEFRL---PRSLSKKKKARVQALIDOLGIRSAKFTVIGD 252
 OY 188 YSLGISTGERRRSIAQLODPKVMLEPDETTGLDCMTANOIVLVLELARNRNIVVL 247
 DB 253 EGRHGVSGGERRRSTIONDIHPDILFLDEPTSGDSTSAVYIKVQRIASGSLVIM 312
 OY 248 THHNPSELFLQDKIAILSFGELIFCGPAEMLDENDCGYPCPEHNPDPFYMDLTSV 307
 DB 313 SHHPSYRIMGLDQLFLSKNGTVYSGSPTHLPQFFSEKHPILPEMENTEERALDLI-- 370
 OY 308 DTOSKEKEITSKRVOMIE-----SAYKSAICHTLNLKNIEMKHLKLP----- 352

DB 371 ---RELEYSTEGTKPLVEPHKOWAKOABSYNNN---NKRNTVSSLKEATIASISRGK 423
 OY 353 -----AVP-EKTDSPGVSKLGLVLRVYRNLVNRNLAVITRLQMLINGLF 399
 DB 424 LVSATNNNSNLPSQOTRANP-FWTEMTYIGKRALNSRQPELLGMRLGAVAVGII 482
 OY 400 LFLFVLVRNSVLKGAIDRVGLLYOFVGAIPYGMILNAVLPFLRAVSDQSDQLYQ 459
 DB 483 LATFTNL-DNSPRGA-QERLG-FFAFMSTFTYCAELPVLEQEXIMREYANAYR 539
 OY 460 KWQMLAYALHVPFSVVAWTFSSVCYWTGLGHPVARFGYSALLAPHLIGEFYTLV 519
 DB 540 RSSYVLSOSIISIPALIVLASFAATFMVAVGIDGANGGFEFFYTLIASFMAGSPVTF 599
 OY 520 LGLVONPNTVNSVVALSTAGVYSGFLRNIOEMPIPKIISTYTPKQYSEI 576
 DB 600 LSGVI--PNWMLGTVVALL--AVFLFGGFFLSRDRIPVYMLFMFHYISLVYYPGVL 655
 OY 577 VNEFYGLNFTGSSNSVSTTNPMCAFQYQIOFIETK 612
 DB 656 QNEF-----QNPTRCFARGVQLDINS 676

RESULT 14

hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text, change 23-Mar-2001
 C:Accession: D96553
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al-
 anse, N.F.; Hughes, B.; Hultzer, L.
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
 C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltl, R.; Marzi
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96553
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-687 <STO>
 A:Cross-references: GB:AE005173; NID:g10092349; PIDN:AA612758.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F5D21.6
 A:Map position: 1
 C:Superfamily: Arabidops thaliana probable ATP-binding cassette protein F12L6.1; A

Query Match 17.58; Score 583.5; DB 2; Length 687;
 Best Local Similarity 28.08; Pred. No. 8e-36;

Matches 167; Conservative 117; Mismatches 239; Indels 73; Gaps 16;

OY 33 PEPHSIGILHASYSGHRVPMMDITSCROQM-----TROILKVSILVESGOICILGSS 88
 DB 13 PPAEIG--KGALV-----WEDLVIVIPNFGGTRRLDGLNHAEGRIAMALMGFS 64
 OY 89 GSKRTLLDMSGRIGRAGTFLGAYVNGRALRREOFODCSYVLOSDTLISSTVRETI 148
 DB 65 GSKSTLDSIAGRLARNVIMTGNLNLNGKARLD--YGLVAVTQEDILMGTLVRETI 122
 OY 149 HTALLAIRGNPSFOKK-----VEAVMAELSLSHVADRILGNYSLGISTGERRVSI 203
 DB 123 TYSAILRL-----SSDLTKEVNDIVSGTITIELGLDCAADRIVGNMHSRGVSGGERRRVSI 178
 OY 204 AAOLADPKVMLEPDETTGLDCMTANOIVLVLELAR--NRNIVLTHHNPSELFLQDK 262
 DB 179 ALFILRPQILFLDEPTSGDLSAFAFYQALRNIRADGRVYSSIHNPSSSEVPALFDD 238
 OY 263 IAILSFGLIFCGTPAEMLDENDCGYPCPEHNPDPFYMDLTSVDTOSKEKEITSKRV 322
 DB 239 LFLSSGELYVFGESKFAVEFPAEAGFPCPKRKNPSDHLRCINSDFIVATLTKSGRI 298

```

QY 323 -----OMISAKKSAICHKTLKNIEMKHLTPMPVPEKTRDSP 362
DB 299 RETPATSDPLMLNLTSEIKARLVEN-YRASYAKSASRIEKLASIGHHEMRKSEA 357
QY 363 GVFSKGLVLRVRRVNLVNRMLAVITRLQWLGFLFFVLRVNSVLKGAIDRVGL 422
DB 358 TWFKQLTKLRKSRVNNCRD-----IGYWSRVIVYIVVSCVGTIFYDVGH 404
QY 423 LYQVVGATP-----YTGMLNANVL--PPVL--RAVSDQSDGLYOKWQMLALAHVL 472
DB 405 SYTSILARVSCGGITFQFMFMSIGSPFTEEKVYKKEKLSGYGVSYIISNYSF 464
QY 473 PFSVVAATMISVSCYWTGLHPEVARFGYSALLAPHLIGELTVLVLGIVQNPINVS 532
DB 465 PFLVALILGISTLYNNMYKFRPGVSHWAFCLNIFSVSIESLMYVAVSLV--PNFLMG 522
QY 533 VVALLSIAG-VIVSGFLRNIOEMPIPF--KIISYFFQKYCSEILVNEFYGLNF 585
DB 523 LITGAGITGITIMTSGEFLRLPDLPKVYMRYPISFMSYGMALQAVKNDFLGEF 578

```

RESULT 15

JC7777

ATP binding cassette transporter, subclass G4 - human

C:Species: Homo sapiens (man)

C>Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 02-Apr-2002

C:Accession: JC7777

R:Engel, T.; Lorkowski, S.; Lueken, A.; Rust, S.; Schlueter, B.; Berger, G.; Cullen, P.; Blochem. Biophys. Res. Commun. 288, 483-488, 2001

A:Title: The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-deriv

A:Reference number: JC7777

A:Contents: Macrophages

A:Accession: JC7777

A:Molecule type: mRNA

A:Residues: 1-646 <ENG>

A:Cross-references: GB:AJ308237

C:Comment: This protein, a novel member of the ATP binding cassette(ABC) half-size trans

C:Genetics:

A:Gene: abcg4

A:Map position: 11q23.3

C:Keywords: macrophage

Query Match 17.4%; Score 578.5; DB 2; Length 646;
 Best Local Similarity 27.1%; Pred. No. 1.9e-35;

Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;

```

QY 33 PERHSGILHASYSHRRVPRMWDITSCQOQTRQILKDVSLVSGQIMCTIGSGSGK 92
DB 54 PRRSANDIEFVELSYSVREGPCW-----RKRGYKTLKLSGKFCRRLIGIMPSGAGK 108
QY 93 TTLDMASGRLAGRAGTFGEVYVNGRALRRBOPOCFVIVOSDTLLSLVRETLHYTA 152
DB 109 STEFNHILAGY--RESGMKQILVNGRPRELRTFRKMSCYIMQDMLPHLTIVLEAMNVA 166
QY 153 LLAIRGNPNGSFOKREAV-----MAELSLSHVADRLIGNYSLSGISTGERRRVSTAAQ 206
DB 167 NLKLEKO-----EVKKELVTEILFALGLMSCHSHTAL-----LSGGQRKLALALE 214
QY 207 LLQDPKVLFDPEPTGLQCMANQIVLVLLVELARRNRIVLTIHOPRSELPQEDKIAL 266
DB 215 LVNPNPVAFEDPSIGLDSASCFQVYSLMKSLAOGRTICTIHOPSAKLFEMFDKLYIL 274
QY 267 SEGELIFCGTAEMLDFENDGYPCEBSNPPDFMDLTSVDYOSKEREIETSKRVOMIE 326
DB 275 SOGQCIFFGVVNTLIPYLGJLRCPTVHNPAFLIEVASG-----EYGDLPMLF 325
QY 337 SAYKSAICHKTLKNIEMKHLTPMPVPEKTRDSP-----GVFSKGLVLRVRRV 378
DB 326 RAV--QNGICAMAEKSSPEKNEVPAPCPPEVDPIESHFTATSTLTQFCILFRKTFPLS 384
QY 379 LVNRKLAVITRLQWLGFLFFVLRVNSVLKGAIDRVGLYQVVGATPYTGMLNA 438

```

```

DB 385 ILRDTVLHLEFNSHVIVGVLIGLILHIGDDAK--VENNTGCLFSSMLFLHFAALMPT 442
QY 439 VNLFPVLRVAVSDQSDGLYOKWQMLALAHVLFPFSVATMTIFSSVCYWTGLHPEVAR 498
DB 443 VLTFPLMAVFMREHLNMYSLKAYVLAKTMDADVPQVVCQVYVCSIVYMTQGPATFSR 502
QY 499 FGYSALLAPHLIGELTVLVLGIVQNPINVSVALLSIAGLVSGFLRNIOEMPI 558
DB 503 FLFESALATATATVAQSLGL-LIGMAASNSQVATFVGVPVVAIPVLFSGFEVSFKTIPT 561
QY 559 KTIISYFFQKYCSEILVNEFYGL--NFTCGSSNVSVTNMPCAFTQIOPIEKTCPG 615
DB 562 LQWSSYLSYVRVGEVYIL-ITGMERGDLTC-----LEERCPRREP-QSILRALDV 611
QY 616 ATSRFTMNLILYSFIPALVILGIYVFKIR 645
DB 612 EDKLYMDFLVLGIFLALRLRLAVLVLR 641

```

Search completed: June 11, 2003, 09:13:51
 Job time : 25 secs

Mon Jul 28 09:43:29 2003

us-09-989-981a-6.rspst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:08:12 ; Search time 39 Seconds

(Without alignments)
3439.402 Million cell updates/sec

Title: US-09-989-981A-6

Perfect score: 3326

Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:REMBL_21.*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	21.2	673	11	Q8R543
2	680.5	20.5	655	4	Q8R7A8
3	672.5	20.2	655	4	Q961A6
4	662	19.9	801	5	Q8R7A8
5	660	19.8	657	11	Q9R004
6	649.5	19.5	725	10	Q9M3D6
7	628	18.9	691	10	Q8RWT9
8	622	18.7	668	10	Q9ARU4
9	618.5	18.6	672	10	Q9ARU4
10	617	18.6	727	10	Q9FNB5
11	615	18.5	692	5	P91892
12	614.5	18.5	703	10	Q8RKN0
13	614	18.5	594	10	Q9LJC3
14	614	18.5	720	10	Q9M2V7
15	610.5	18.4	725	10	Q9ZU35
16	610.5	18.4	725	10	Q9ASR9

17	610	18.3	679	5	Q9BH97
18	608	18.3	708	10	Q9M2V5
19	602.5	18.1	654	10	Q9L1W2
20	600.5	18.1	670	5	Q77423
21	600	18.0	602	5	Q9VC15
22	597	17.9	590	10	Q9MAH4
23	595.5	17.9	658	5	Q16574
24	595.5	17.9	687	5	Q94960
25	595.5	17.9	785	4	Q96176
26	592	17.8	740	5	P90746
27	591.5	17.8	646	10	Q80946
28	589.5	17.7	646	10	Q966R7
29	588.5	17.7	648	10	Q966W5
30	587.5	17.7	627	11	Q91WA3
31	587.5	17.7	646	11	Q8VBS9
32	584	17.6	755	10	Q9ZUT0
33	583.5	17.5	687	10	Q9C8K2
34	581.5	17.5	687	5	Q9NH94
35	580.5	17.5	669	5	Q8RRE2
36	578	17.4	705	10	Q9L1W1
37	575.5	17.3	649	10	Q9S1T6
38	575.5	17.3	666	11	Q9EPG9
39	573.5	17.2	669	5	Q8MRRL
40	572	17.2	739	10	Q9LFC8
41	571.5	17.2	609	10	Q9C8W6
42	571.5	17.2	662	10	Q949Y4
43	571	17.2	609	5	Q9YV04
44	567	17.0	626	5	Q8T684
45	565	17.0	1528	5	Q8T677

ALIGNMENTS

RESULT 1
ID Q8R543 PRELIMINARY: PRT: 673 AA.
AC Q8R543;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Steroidin 2.
GN ABCG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;
RT "Molecular cloning, genomic structure and characterization of novel
RT mouse head-to-head tandem ABC transporters";
RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF351811; AAL82898.1;
DR EMBL: AF351799; AAL82898.1; JOINED.
DR EMBL: AF351800; AAL82898.1; JOINED.
DR EMBL: AF351801; AAL82898.1; JOINED.
DR EMBL: AF351802; AAL82898.1; JOINED.
DR EMBL: AF351803; AAL82898.1; JOINED.
DR EMBL: AF351804; AAL82898.1; JOINED.
DR EMBL: AF351805; AAL82898.1; JOINED.
DR EMBL: AF351807; AAL82898.1; JOINED.
DR EMBL: AF351808; AAL82898.1; JOINED.
DR EMBL: AF351809; AAL82898.1; JOINED.
DR EMBL: AF351810; AAL82898.1; JOINED.
SQ SEQUENCE 673 AA; 76008 MW; FA08340445DE259C CRC64;
Query Match 21.2%; Score 704; DB 11; Length 673;
Best Local Similarity 28.7%; Pred. No. 5.8e-43;
Matches 195; Conservative 130; Mismatches 261; Indels 94; Gaps 17;
11 GSMGLQVNRGSSSLEGAAPATAPPHSHGILHASYSVSHRYR-----PMD-I 57

```

Db      14 GVLADASQGLDSDL-----ESSESNSTLYFTYSGSNTLEVDLTGYVDIASQVWFEOQL 69
Oy      58 TSCROQWTRQI-----LKDVSLYVESGOIMCIGSSGSKTTLIDMAGRLRAGTF 109
Db      70 ADFKIPMRSHSSQDCELGIRNLKRVNSGQMLAITGSSGCRASLDVYITR-CHGCKM 128
Oy      110 -LGEVYVNGRALRBOQDFCSYVLQSDPLTSLSTLVRETLHYTALLAI-RGNGPSFQK 167
Db      129 KSGQIMINGQSTPOLVKRCVAHVHQHQLPLNLTVEFLATIAQMLPRTSQQRDR 188
Oy      168 VAVVAEELSHVADRLGNSTLGGISTGERRRVIAAQLDPPVMLEFDEPTGLDCMT 227
Db      189 VEDVAIEALRLQCAATRGNTYVRGSGERRRVSIGVLMNPGLILIDEPTSGLDSET 248
Oy      228 ANQIYVLELARRRRIYVLTIHOPRSELPOLFQIALISFGLFCGTPAAMDLEFND 287
Db      249 AHNLTTLRLAKGRVLVLSLQPRSDIFLELDVLTMTSTPYILGAQOQVQYFISI 308
Oy      288 GYPCPEHNPDEYMDLTSVDTOSKERELETSKRVOMIESAYKKA-----ICHKTLKN 341
Db      309 GHPCPRYNADPFYDLTSDIRSKEREVAAYEKQSLALELEKVGDFDLKMAEKE 368
Oy      342 IERMAHLKTPVAPFKTKDS-----PGVFSKGLVLRVTRNLVRNKLAIVTRLLON 393
Db      369 LNTSTHTVSLTL-----TQDTCGTAVELPQMEQFSTLIHQISNDFRDLPTLLIHGSEA 424
Oy      394 LIMGLELFEVLRVNSVNLKGIQDRVGLVQFVATPYTGMLNVLNPEVLRASDDES 453
Db      425 CLMSLITIGFLYGHAKOL--SEMDTALFLMIGALIPFNVLIDVYSSCHSERSMLYEL 482
Oy      454 QDGLYQKQOMLAVLHVLPESVVAITMIFSSVCYTLLGLHPEVANRFGFSALLPHLIG 513
Db      483 EDGLTAGYFFAKILGELPEHCAYVIYAMPIYMLNLRP-----VPEFL 529
Oy      514 EELTVILGIVONPPIVNSVALLS-----IAGVLYGS-----GFLRNIQEMDIPKRI 561
Db      530 LHELVLMVLVFCRCRNMALASAMLTPTFMSSFCNALNSFTLTGFMINDMLIVIAM 589
Oy      582 ISYFFOKYCEELVAVNEFYGLNFT--CGSSNVSTYTPMCAFTGIGQIEKTCGATSR 619
Db      590 ISKLSFLKRCPSGLMOIOPNGHLTYTOIGNFTFSILGDTM-----ISA 632
Oy      620 FTMNPLIYSPFIPALVILGI 639
Db      633 MDLNSHPLYAYI-LIYIGI 650

RESULT 2
Oy      967A8 PRELIMINARY; PRT; 655 AA.
AC      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DB      ATP-binding cassette superfamily G (White) member 2.
GN      ABCG2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN      NCB1_Taxid-9606;
RP      SEQUENCE FROM N.A.

RT      MEDLINE-21201983; PubMed-11306452;
RA      Komatsu H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,
RA      Arakawa H., Nishimura S.;
RT      Identification of breast cancer resistant protein/mitoxantone
RT      resistance/placenta-specific, ATP-binding cassette transporters as a
RT      transporter of NB-506 and J-107088, topoisomerase I inhibitors with an
RT      indolocarbazole structure.
RL      Cancer Res. 61:2827-2832(2001).
DR      EMBL; AB051855; BAB46933.1;
DR      InterPro; IPR003439; ABC_transportr.

```

```

DR      InterPro; IPR003880; Pantane_attach.
DR      InterPro; IPR003580; Prochachykinin.
DR      Pfam; PF00005; ABC_tran.1.
DR      ProDom; PD000006; ABC_transportr.1.
DR      SMART; SM00203; 7K; 1.
DR      PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
DR      ATP-binding.
DR      KX
SQ      SEQUENCE 655 AA; 72314 MW; ABAF6B96034C5A8 CRC64;

Query Match      20.5%; Score 680.5; DB 4; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.9e-41;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Oy      21 SSSLEGAPATP---EPISILNLSVSVSRVPMMDITSCROQWTRQIKDYSIYE 77
Db      13 SGNNGNPFATASNDLKAFTBEGAVLSFHNICRYKLKSGFLPCRPRVEKILLSNIGTK 72
Oy      78 SQOIMCIGSSGSGRTTLIDAMSGRLGRAGTFLEGYVNGRALRBOQDFCSYVLQSDPT 137
Db      73 PG-LNALIGPTGGKSSLLDVLAAKRDPSG-LSGDVLING-APRANFNCSGYVQDDV 129
Oy      138 LLSLTVRETLHYTALLAIRGNPG-SFOKKVAVMAELSHVADRLIGNSTLGGISTG 196
Db      130 VMGTLTVRENLQFSALRLATMTNHEKNERINRYOELGLDKVADSKVGTQFIRGVS 189
Oy      197 ERRRVIAAQLDQDKVMLEFDEPTGLDCMTANQIYVLELARRRRIYVLTIHOPRSEL 256
Db      190 EKKRISIGMELTDSILFLEDEPTGLDSTANAVALLKRSKOGRTTFSIHOPRSI 249
Oy      257 FOLFPIKIALISFGLFCGTPAEMLEFNDGYPCEPHSNPDPFYMDLTSVTO-----SK 312
Db      250 FKLFDLSLTLASGRIMFHPGPAQELGYFESAGYHCEAVNNPADFLLIDNGSTVALNR 309
Oy      313 ERE-----IETSKR-----VOMIESAYKKAICHT-----LKNIERKHLKTLPMVPE 356
Db      310 EEDFKRATETIEPSKODKPLIEKLAIVYSSPYFKEKAKELHOLSGGEKKKTKVREISY 369
Oy      357 RTKDSPGVFSKGLVLRVTRNLVRNKLAIVTRLLONLIMGFL--LEFVLRVNSVNLKGI 414
Db      370 TT-----SCHQLRWMSKSKFKLLNPPQASIAQIIVYVVLGLVIAITFGLANDST 421
Oy      415 AIQDRVGLLYQFVAPPTGYMLNVLNPEVLRASDQDGLYQKQOMLAVL-HVLP 473
Db      422 GIONRAGVLF-FLTNGCFSSVSAVELFVVEKKLTHIHYISGYVSSYFGKLKSLDLP 480
Oy      474 FSVVATMIFSSVCYTLLGLHPEVANRFGFSALLPHLIGELTVILGIVONPPIVNSV 533
Db      481 MEMLPSTILFTCIYVMLGLKPKRDAFVYMMFTLM--VAVSASSMALAIAAGOSVSA 537
Oy      534 VALLSIAGV--LVGSGFLRNIQEMDIPKRIISYFFOKYCEELVAVNEFYGLNFTCGSSN 591
Db      538 TLMATICFVPMATFSGLVNLTITIASMSLWIOYFISIPRYGFLALQHNELGONFCPG 594
Oy      592 VSVTTNPMCAFTGIGQIEKTCGP 615
Db      595 INATGNPCNVA-----TCTG 610

RESULT 3
Oy      967A8 PRELIMINARY; PRT; 655 AA.
AC      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DB      ABC transporter ABCG2.
GN      ABCG2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN      NCB1_Taxid-9606;
RP      SEQUENCE FROM N.A.

```


OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FVB: TISSUE=LIVER;
 RX MEDLINE=99413474; PubMed=10485464;
 RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
 RT "The mouse Bcrp/Mxr/Abcp gene: amplification and overexpression in
 cell lines selected for resistance to topotecan, mitoxantrone, or
 doxorubicin";
 RL Cancer Res. 59:4237-4241(1999).
 DR EMBL: AF140218; AAD54216.1; -;
 DR MGI: MGI:1347061; Abcg2.
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR003880; Pantane_attach.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transportr.1.
 DR PROSITE: PS00012; PHOSPHOTRANSFERASE;
 SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 19.8%; Score 660; DB 11; Length 657;
 Best Local Similarity 28.0%; Pred. No. 9e-40;
 Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

QY 13 MGLVNRGSSSLGAPATAPERSIGILNASYSVSHRVRPMDITSCROQWTOIILKDV 72
 Db 12 MSQNNNGQLPMMNSRAVTLAEGDVLSEFHITTYRV--KVKSGFLV--RKVEKEIISDI 66
 QY 73 SLVYESQIMCIIIGSSSGKTTLLDAMSGRLGRGTGLGEVYVNGRALRQPODCSYV 132
 Db 67 NGIKPKP-LNALILCLPTGGKSSLLDVLAAKRDPRG-LSGDVLNG-APQAHFRCSSGYV 123
 QY 133 LQSDTLSSLVRETLVYTLALAIRGNPSFQ-----KKEAVMAELSHVADRILGN 187
 Db 124 VQDDVVGTLVVRNLQPSALRL-----PTMKHMKERNRNTIKELGKLVADSKVGT 179
 QY 188 YSLGISTGERRRVSIAQLQDPRVLFDEPTGLDQMTANOIVLVLELARRRIVVL 247
 Db 180 QFINGISGGERKRTSIGMELITDPSILFDEPTGLDSTFANAVILLKLRMSKGRITIF 239
 QY 248 TIHPRSELPFLPKIALISFGLIFGCTPRMELDFNDQYPCPEHSNPFDFMDLTSV 307
 Db 240 SIHPKRSIFKFLPSLTLASGLKLVFHPQAKALEYRASAGYHCEPYNPAPDFLDVING 299
 QY 308 DTOS-----KERLETSKR-----VOMISAKKSAICKTLKNIERRMKHLTLP 352
 Db 300 DSSAVMLNBERQDEANKTEPEKGVLENLSEFYINSALVGETRAELDQ----- 352
 QY 353 MWPKTKDSPGVESKLV-----LIRVTRNLVRRKLAIVITRLQNL 394
 Db 353 -----PGAOEKKGTSAPKEPEVYTSFCHQLRMIAKRSFKMLGNPQASVQALVTV 403
 QY 395 IMGLFL--LPEVLVRSNVLKGATODRVGLLYOVGAPRYTGMLNANVLFPVLAAYSDQ 452
 Db 404 ILGLILGAIYFDLKYDA---AGMONRAGVLE-PLTNQCSVSSAVLEFVEKKLPIHE 458
 QY 453 SQDLGYKQWOMLALVAL-HVLPFSVAVMIFSSVCYVTLGLHPVARFGYSALLAPHL 511
 Db 459 YISGYVRSYSFEKGVMSDLLPMRFLPSVIFTCLLYLMGLKAKTVDAFFIMFLLI--M 515
 QY 512 IGEFLTLVLYLGYONPVIWVSVALSLIAGV--LVSGGFLRNIOEMDIPKXIISYFQK 569
 Db 516 VAYTASSMALAIAGOSVSAVALTIAEFVFMILFSGLLVNLRTIGTPMLSMILQYFSIPR 575
 QY 570 YCSLIVVNEFYGLNFGSSSNVSVTNPMCAFTQGIQFIKTCRG 615
 Db 576 YGFLALDYNEFLGEGCPG--FNVTDNSTCVNSIAT-----CTG 612

RESULT 6
 ID 09M3D6 PRELIMINARY; PRT: 725 AA.
 AC 09M3D6;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ABC transporter-like protein (Putative ABC transporter protein).
 GN T26112.10 OR AT3G55130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Monfort A., Casacuberta E., Pulgomenec P., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.H., Tang C., Toriumi M.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Natsuka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T26112.10/AT3G55130 (GI:7019646).";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132954; CAB5747.1; -;
 DR EMBL: AY045932; AK76606.1; -;
 DR EMBL: AY079387; AL85118.1; -;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transportr.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transportr.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN.1.
 DR KATP-binding.
 DR KATP-binding.
 SQ SEQUENCE 725 AA; 80656 MW; 790C535A7929C16 CRC64;

Query Match 19.5%; Score 649.5; DB 10; Length 725;
 Best Local Similarity 29.4%; Pred. No. 6e-39;
 Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

QY 33 PEPHSLILNASYSVSHRVRPMDITSCROQWTOIILKDVSLVYESQIMCIIIGSSSGK 92
 Db 68 PVPYLVANFNLIQYVTLARR-----FGSRQNGKVTLLDDVSGASDSDIILAVGASAGK 123
 QY 93 TLLDAMSGRLGRGTGLGEVYVNG-RALRREDQFQDCFSYVLSQDTLSSLTVAETLHYT 151
 Db 124 STLDALAGRAVE-GSLRGSVTLNGEVLKSLRLKLVISAVYMDODLFLPLUYKETLMPA 182
 QY 152 ALTAIRRG-NPGSOKKVEAVMAELSHVADRILGRTSLAGISTGERRRVSIAQLQD 210
 Db 183 SEFLRPSLSKSKMKMEVEALLDQLGRNANVTYIGEGHGVSGGERRRVSIGIDIHD 242
 QY 211 PKVMLFDEPTGLDQMTANOIVLVLELARRRIVTLTIHQPSSELPQLDKTALISFGE 270
 Db 243 PIVLFLDEPTSGLDSTAFANVYVLRKIAOSGSIVINSHQPSARIYELDLRLITLSRKG 302
 QY 271 LIFCGTFAEMLDFNDQGYPCPEHSNPFDFMDLTSVDTSQSKRELETS----- 319

Db 303 SVFNGSPASLPGFSDGRIPEKENISEFALDIV-----RELEGSNEGTRALVDEN 354
 QY 320 -----KRVOMIESAYK-----KSAICHKTL--KNIEEMKILKLPMPFETKD 360
 Db 355 EKMOONKISLISQSPQTNKLDQDRLSKENAINASRGKLVSSGSSNPTSMETVSSYA 414
 QY 361 SPGVESKLGVLRRVYRNRLVKNLAVITRLQNLIMGLFLFVLRYRSNVLAQAIDRV 420
 Db 415 NPSLFEFF-ILAKRYMNMIMPELVGTRIATVMTGC--LLAVYMKLDHTPPGA-QERL 471
 QY 421 GLLYQFVGATPYTGMLNAVLFPVLRAVSQESQDLDYQKQOMLALYALHVPFVATM 480
 Db 472 -TLFAFVPTMFCCLDNPVPIQERITFLRETHNAVTRSSVYISHSLPQLLAPSL 530
 QY 481 ISSVCYWTGLHPEVAREGYSFSAALLAPHLIGEFLTVLGLGVQNPVINS-VVALSTI 539
 Db 531 VESATFMTVGLSGLEGEFVFCILYASFSGSSVYTFISGV--PNIMLCYMSITYL 588
 QY 540 AGVLGSGFLRMIOEMPIPKIISTFPOKYCEILVNEVYGLNFTCGSSNVYTTNM 599
 Db 569 AYCLLSGTYVRDRIPEYTWTFHYISILKYEAVALINEF-----DDPS 633
 QY 600 CAFTGQIPIERTCGATS 618
 Db 634 RCFVRCGVFPDSTLLGVS 652

RESULT 7

Q8RM19 PRELIMINARY: PRT: 691 AA.
 ID Q8RM19; 01-JUN-2002 (TREMBlrel. 21, Created)
 AC Q8RM19; 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE Hypothetical. 77.2 kDa protein.
 GN A7G321090.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carolin P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shino P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY093054; AAM13053.1;
 KW Hypothetical protein.
 SQ SEQUENCE 691 AA: 77219 MW: CEA73CC0B40D7E9 CRC64;

Query Match

Best Local Similarity 18.9%; Score 628; DB 10; Length 691;
 Matches 173; Conservative 123; Mismatches 225; Indels 94; Gaps 17;

QY 25 LEGAPATAPE-PHISGILHASYSVSHRRPMDITSCROO---TROILDVSLYVSG 79
 Db 3 LEGSSSGRQRPSTLEMSRGAVLA-----WEDLVVVPNSDPTTRLRLQRLNGYAE 56
 QY 80 QIMCIISSSGSGKTTLLDAMSGRLGRAGTFLGEYVNGRARRQFODCSYVLQSDTL 139
 Db 57 RIMAIMPGSSGKSTLDSLAGRLARNVMTGNLLNGKARLD--VGLVAVYQEDVLL 114
 QY 140 SSVLTRETLHTALAIRNGPSFQK-----VEAVMAEISHVADRILIGNSLQGIS 194
 Db 115 GTLVRETIYSALRL-----PSDMKEEYSDIVEGITIELGLDCCSDRYIGMHWARG 170
 QY 195 TGERRRSIAAQLLODPKVALPDEPTGLDCMTANOIVLLVELARRNRIVLTIHQPRS 254
 Db 171 GGRKRVSTALETLRPOILFDEPTSGDLSASAFVQLARNIADGRIVISSVQPPSS 230

QY 255 ELRQFPKIALISNGELIFGCTPAEMLDFFNDCCYPCPEHNSNPPDFMYDLTSVDTOSKER 314
 Db 231 EVRALFDLFLSSGESVYGEAKSAVEFAESGFCPCPKRNPSDHLRLCINSFDTVTA 290
 QY 315 EIETSKRVQ-----MIESAKYKSAICHKTLKNIEEMKILKLPMPV 354
 Db 291 TLKSGRIQETPATSDPMLNATSVIKARLEN-YKRSKAKSARIRRELINEGLEME 349
 QY 355 PFTKSPGVESKLGVLRRVYRNRLVKNLAVITRLQNLIMGLFLFVLRYRSNVLA 414
 Db 350 IRKSGEATW-MKQFLRTLRSFIMCRDGVYTRITISYIVSI----- 392
 QY 415 AIODRGLLYQFVGATPYTGMLNAVL-----FPVL---RAVGDOSQDS 456
 Db 393 ---SVGTLFVYGVGS-YYSILARVSCGGEFTGFMFMISGIFSPLEEKYFKERLSG 447
 QY 457 LYQKQOMLALYALHVPFVSVAADMIFSSCYTGLHPEVAREGYSALLAPHLIGEFL 516
 Db 448 YGVSVYILSNISSEPFVLAISVITGITTYMLVFERPQFSHYAFCLNIFPSVYTESL 507
 QY 517 TLVLGIVQNPVINSVVALLSIAG-VLVGSGFLNIOEMPIPKI-----ISTFPOKY 570
 Db 508 MMYVASVY--PFLMGLITGAGLIGITIMTSGFFRLDLP---KIFVRPVSYSTYSGSW 562
 QY 571 CSEILVNEVYGLNF 585
 Db 563 AIOGGYKNDFLGLEF 577

RESULT 8

Q9ARU4 PRELIMINARY: PRT: 668 AA.
 ID Q9ARU4; 01-JUN-2001 (TREMBlrel. 17, Created)
 AC Q9ARU4; 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE Putative ABC transporter.
 GN P0445D12.3.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eurytharoidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0445D12.3";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1. SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AP003046; BAB40032.1;
 DR InterPro: IPR003593; AAA_Artpase.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 668 AA: 73368 MW: D1875B8C75B0F3B2 CRC64;

Query Match

Best Local Similarity 18.7%; Score 622; DB 10; Length 668;
 Matches 171; Conservative 100; Mismatches 245; Indels 36; Gaps 12;

QY 55 WDTISC-----RQMTROIILKDVSLVYEQIMCIISSSGKTTLLDAMSGRLGRAGT- 108
 Db 58 WATITALKNRKRDVARFLLSNASGEAKSGRLALAMPSSGKTTLLANVLAGQLTASP 117
 QY 109 -FLGEYVNGRARRRQFODCSYVLQSDTLSSLYVRETLHTALAIRG--NPGSFQK 166
 Db 118 HLGSLYINRPTISEGCK--LAVVROEDLFFSGLVRETS- 230

Oy		167	KVEAVMAELSLSHVADRLIGNTSLSGISTGERRRRVSAIAQLLDDPKMLFDEPTTGIDCM	226
Db		176	YVNDELFRLLGNACADSIYGAOKVAGISGGCKRRLSLAELLASPIIFADEPTTGLDNF	235
Oy		227	TANQIVLVLLVARNRNRIYVLTIHQPRSEFLFDLKIALISFGCLIFCGTPA--EMLDPF	284
Db		236	QAEKMEWTLROLAEEGHVTCISIHOPRGSVYGFDDIYLLSEGEVIYM-GPAKEEPLLYF	294
Oy		285	NDCGPRCEHNSPDPFYMDLTSPVDOSKRELETISRVO-MIESAKKSN--CHTKLN	341
Db		295	ASLGHCEDPHNPAPFEFLDLSDIVDSASVSRSKRRIENLIEESNKVAITESNSLTN	354
Oy		342	IERNKHLTPLVPVPKTDSPGVSKLGYLRRV-----TRNLVRNKLAIVTRLQN	393
Db		355	PBGSEFSFKL-IQSTTKHRBWMNQFLLEFRAMQAARDGPTNNVRAKMSASA----	409
Oy		394	LIMGLELLEFYLVRVSNVKGAIDORVGILLYOFVGATPYTGMINAVNLFPPVLRVADSQS	453
Db		410	-----IIFGSYFMHGKTQTSLDORMGLLOVTAINTMAALTKTVGVEPKERAIVDRER	463
Oy		454	ODGLCTOKNMMLAYLVLPFSVNAFTMISSVCYMTGLGHPARARGYSALLAHHLIC	513
Db		464	AKGSTALGPYSSKLLAEIPIGAAPPLLGSILYPMSKLPPFSRAKFCGTIVESFAA	523
Oy		514	EFLTYLLGIYONPNIVNSVVALLSIAGLVSGFLRNIOEMPRIKITSYTFPOKYCSE	573
Db		524	SAMGLTVCAMAPTTEAAMAALGPSLMTVFIFYFG-GIYVNDPNFTVIRMLPKVSLINKAFQ	582
Oy		574	ILVYNFFYGLNF	585
Db		583	GLCINEFKGLDF	594

RESULT 9

09LI82		PRELIMINARY:	PRT;	672 AA.
09LI82				
AC				
DT		01-OCT-2000 (TREMBLrel. 15, Created)		
DT		01-OCT-2000 (TREMBLrel. 15, last sequence update)		
DT		01-JUN-2002 (TREMBLrel. 21, last annotation update)		
DE		ABC transporter-like protein.		
OS		Arabidopsis thaliana (Mouse-ear cress).		
OC		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
CC		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OX		eurosidia II; Brassicales; Brassicaceae; Arabidopsids.		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN-COLDBAIR;		
RA		Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;		
RL		Submitted (MAR-2000) to the EMBL/Genbank/DDay databases.		
RN		[2]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN-COLDBAIR;		
RX		MEDLINE-20363099; PubMed-10907853;		
RA		Nakamura Y.;		
RT		*Structural analysis of Arabidopsis thaliana chromosome 3. II.		
RT		Sequence features of the regions of 4,251,695 bp covered by ninety pl,		
RT		TAC and BAC clones.*;		
RL		DNA Res. 7:217-221(2000).		
DR		EMBL: AP001313; BAB03081.1;		
DR		InterPro: IPR003593; AAA_ATPase.		
DR		InterPro: IPR003439; ABC_transporter.		
DR		InterPro: IPR003880; Pnante_attach.		
DR		Pfam: PF00005; ABC_tran; 1.		
DR		ProDom: PD000006; ABC_transport; 1.		
DR		SMART: SM00382; AAA; 1.		
DR		PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.		
DR		PROSITE: PS00012; PHOSPHOTANETHEINE; UNKNOWN_1.		
KW		ATP-binding.		
SEQ		SEQUENCE 672 AA: 75269 MW: 208BD99215600135 CRC64;		

Query Match	18.6%	Score 618.5	DB 10	Length 672
Best Local Similarity	32.1%	Pred. No. 9.8e-37		
Matches	182	Conservative 100	Mismatches 206	Indels 79
			Gaps 17	
QY	66	ROLIKDVSIVYESQIMCIIGSSSGKTTLDAMSGRGRAGTFLGEYVNGR----	ALR 121	
DB	97	RLVTKCVSGVIGKPELLAMLPGSSGKTTLTALAGRL--OGKLSGVSYNGEFTSVK	154	
QY	122	REGQDCPSVYLOSDTLSSLTVEETLHYTALLAIRGNPPSPCK-----VEAVMELS	176	
DB	155	RKT-----GEYQDVVLPHLTMEETLYTALLRLR---PRLTRKEKLEOVENVSDLG	205	
QY	177	LSHAVDRLLIGNYSLGISSTGERRRVSIQAOLLODPKVALFEPTTGDCMTANOIVLLV	236	
DB	206	LTRCNASTYIGGGLRGISGGRKRVSIQGEVLVNPSSLILDEPFGSDSTNAITVALR	265	
QY	237	ELARNRIVLTLIHQPSSELFOLDKTAIISFGLIEFCGTDAEMIDFNDGCV-PCPEHS	295	
DB	266	SLANGGRVYVTTIHQPSRLYRMPDKVLYLSEGCPIYSGDSGRVMEYFGSIGYOPGSSFV	325	
QY	296	NPPFVMDLTVDVQSKER--EIEFSKRVQME-----SAYKRSALCHTKMKIE	343	
DB	326	NPADVDLDLAGCITSDTKQYQIETNGRLDRLEONSYKQSLISSYKN--LYPLKE--	381	
QY	344	RMKHLKTLPMVPFKTKSPGV-----FSKLGVLLRLRVNRLVNRKLAIVT	388	
DB	382	--EVSRTFP-----QDOTNARLRKKATITNRPMPSMMQGFVLK---RGLKSHSESFS	430	
QY	389	RLNLNLINGFLFLEFVLRVNSVLKGAIDRVGLLYQFVGATPYTGMLNAVNLFPVLAV	448	
DB	431	GLRIFMVVSLSLGLLMMHSRV--AHLQDVGLLFFSIFMGFFPLNALFTFPQERPM	488	
QY	449	SDQESQGLYOKWQMLALVHLVLPFSVVAITMISSTVCTWTLGLHPEVARGYSALLA	508	
DB	489	LIKESSTGYSRLSYTARTGDIPEMLILPTIYVTVYMGKGKPSLTTFITMTLMTLY	548	
QY	509	PHLGEFTLVLLGIVONPNIVNSVVALSTAGLVGSGFLRNLOEMPPIPKTISYTFEQ	568	
DB	549	NVLVAQGGGLGAILMDAKRAATISVLMVFLLAGSY--IHLHGTAMLKYSFS	605	
QY	569	KYCSIELLVNFEYGLNF-----CGS 589		
DB	606	HYCKLAV-----GVQYTWDEYVEGGS 627		
RESULT 10				
Q9FNBS PRELIMINARY; PRT: 727 AA.				
ID	Q9FNBS			
AC	Q9FNBS			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DS	ABC transporter-like protein.			
DE	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:			
OC	Spermatophyta: Magnoliales: eudicotyledons: core eudicots: Rosidae:			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RX	MEDLINE=98069011; PubMed=9405937;			
RA	Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. II.			
RT	Sequence features of the regions of 1,044,062 bp covered by thirteen			
RL	physically assigned P1 clones.";			
RL	DNA Res. 4:291-300(1997).			
DR	EMBL, AB006704; BAB08684.1. -			
DR	InterPro: IPR003593; AAA_Artpase.			
DR	InterPro: IPR003439; ABC_transportr.			
DR	Pfam: PF00005; ABC_tran. 1.			
DR	ProDom: PD000006; ABC_transportr. 1.			

DE Putative ABC transporter protein.

GN ATG17840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.
RT "Arabidopsis Full Length cDNA Clones";
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY080792; AAL87274.1; - to the EMBL/Genbank/DBJ databases.
SQ SEQUENCE 703 AA; 78412 MW; 8085EE940AC39003 CRC64;

Query Match 18.5%; Score 614.5; DB 10; Length 703;
Best Local Similarity 29.4%; Pred. No. 2e-36;
Matches 170; Conservative 112; Mismatches 233; Indels 63; Gaps 16;

QY 47 VSHRPPWMDIT---SCROQTROIKADVSLVESGGIMCIGSSGSKTLLDMSGR 103
DB 46 VSARLTWQDLVAVYTMGDGTQNVLEGLTGYAEPSGLTALPGSGSKSTMALASRL 104
QY 104 GRACFFLEGVVNGALRREPQDQCFSTVYLOSPTLLSLYRETIHTALAIRGNFGS 163
DB 105 AANAFSLGTVLNGR--KRLSPGTAAVYVDDNGLIGTLVRETIWYSARVLT---PDK 158
QY 164 FQK-----VEAVMAELSHVADRLIGNYSIGISTGERRRVSIAQLDLPKYLDE 218
DB 159 MLRSKRALVERTIEMGLDCAQVYIGNMHLRGISGSEKRRVSTALEILMRPLFDE 218
QY 219 PTTGIDCATNOIYVLLVELARRNIVLTJHOPSELPFLDKTALISFELJFCGTPA 278
DB 219 PTTGIDCATNOIYVLLVELARRNIVLTJHOPSELPFLDKTALISFELJFCGTPA 278
QY 279 EMLDFNDCGXPCEPHSNPDEYMDLTSSV-----TQSERIETESK 320
DB 279 DATEFFAAGFCPCPLRNPSDFLNCINSDPKVATILKSNKLEFASDDPLEKITTAE 338
QY 321 RVOMIESAYKKSALCHTKLNIERMKHLKTLPMVEFKTDSPGVSKGLVLRVTNLY 380
DB 339 AIRLLVDYHTSDYYTAKAVEEISQFKTIL-----DSGG--SQASFLLQ--TYTLT 388
QY 381 RNLAVITRLLONLIMGLFLFVLVRVSNVLCGAIDRVGLLYQEV-----GATPYTGM 435
DB 389 KRSPINMSR---DFGYMLRLILYILYVCIIGIYLVNVTGSYSAIILRSCASFFVGF 443
QY 436 LNAVNL--FPVL---RAVSDQSDGLVOKWOMLAVALHLPVSVAITMIFSSCYTL 490
DB 444 VTFPSIGCFSPFVEDMKVFORERLNGHGVAAFYANTLSATPILMTITPISGTCYMW 503
QY 491 GLHPEVAFGYFSALLAPHLIGEFLLVLLGIYONPNIIVSVALLSIAGV-LVSGFL 549
DB 504 GLHGFHYLFEFLVCLVASYVESLIMAINSIY--PRLMGIIIGAGIOGIFMLVSGFF 561
QY 550 RNIDEMPIPF--KIISTFTFOKYSEILVNVNRYGLNF 585
DB 562 RLPPNDIRKPFWRYPMSYSIFHFVALOGYOYNDRLGLTF 599

RESULT 13

09LJC3 PRELIMINARY; PRT; 594 AA.
AC 09LJC3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE ABC transporter-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RA Nakamura Y.;
RX MEDLINE-20363099; PubMed-10907853;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AP000604; BAB01452.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transporter.1.
DR SMART: SM00382; AAA.1.
DR PROSITE: PS00211; ABC_TRANSPORTER.1.
KW ATP-binding; Transport.
SQ SEQUENCE 594 AA; 65939 MW; 5BD2F5300DB36CA8 CRC64;

Query Match 18.5%; Score 614; DB 10; Length 594;
Best Local Similarity 28.0%; Pred. No. 1.8e-36;
Matches 173; Conservative 123; Mismatches 229; Indels 92; Gaps 17;

QY 25 LEGAPATAPE-PSHIGILHASYSVSHRVPMDITSCROQ---TROILDVSLYVSG 79
DB 3 LEGSSGRRQIPSKLEMSRGVLA-----WEDLVVLPNNSDGTTRLLRLINQYAPG 56
QY 80 QIMCLSSGSGSKTLLDMSGRGRATFLGEYVYVNGRALRREPQDQCFSTVYLOSPTLL 139
DB 57 RIMATMGSSGSGSKTLLDMSGRGRATFLGEYVYVNGRALRREPQDQCFSTVYLOSPTLL 114
QY 140 SSLYVRETIHTALAIRGNFGSFOK-----VEAVMAELSHVADRLIGNYSIGIS 194
DB 115 GTLVRETIHTYSAHLRL-----PSDKSREVEDIYDGTIMELGLDCCSRVGNHARVYS 170
QY 195 TGERRRVSIANAQLDQPKVMLFDEPTGLDCTANQIYVLLVELARRNIVLTJHOPRS 254
DB 171 GGERRRVSIANAQLDQPKVMLFDEPTGLDCTANQIYVLLVELARRNIVLTJHOPRS 230
QY 255 ELFOFLDIALISGELIFCGTPAEMDFFPNDGCPGCEHNPDPFTYMDLTSVQSKER 314
DB 231 EVFALEFDLFLSSGESYFGEAKSAVEFFAESGFCPKKRNPSDHLRCINSDFDVTA 290
QY 315 ELETSKRVO-----MIESAYKKSALCHTKLNIERMKHLKTLPMV 354
DB 291 TLKSQRQIETPATSDDLMLNATSVIKARLVEN-YKRSKYAKSKSRRELSTNIEGLEME 349
QY 355 PFKTKDSGCVSKGLVLRVTRNLVNRKLAIVITRLLONLIMGLFLFVLVRVSNVLCG 414
DB 350 IRKSEATW--WKQLRITLARFINNCRDVGYYMTRISYIVSI----- 392
QY 415 AIDRVGLLYFGVATPYTGLMNLVNL-----FPVL---RAVSDQSDG 456
DB 393 ---SVGITFYDVAYS-YTSLIHAVSGGFTTGFTMSIGCFSPLEMKVFPYKERLSG 447
QY 457 LYOKWOMLAVALHLPVSVAITMIFSSCYTLGLHPEVAFGYFSALLAPHLIGEFLL 516
DB 448 YGVSVYILSNYISFPFLVAIVITGTITYNLVKFRPGSHVAFCLNIFSVSYESTL 507
QY 517 TLVLGIYONPNIIVSVALLSIAG-VLVSGFLRNIDEMP---IPRITISYTPQKY 570


```

OY 160 NPGSFKVEAVMAELSLSHVADRLIGNVSLGISTGERRRVSIQAOLIDPKVMLFDEP 219
Db 183 SAEERDEYVNMILLKLVSCADSCVDKAVRGISGGEKKRLSLACELIASPSVIFADEP 242
OY 220 TTGLDQCTANOIYVLLVEIARNRRIYVLTJHOPRSELFQLPDKIAILSGELIFGCTPA- 278
Db 243 TTGLDAPQAEKVMETLQKLAODGHVICSIHQPRGSYAKFDDIYVLLTEGTLVYAG-PAG 301
OY 279 -EMLDFFNDGYPCEPHSNPFDYMDLTSVDTSQEREIETSKRVOMIESAY--KKSATC 335
Db 302 KEPLTYFCNFGFLCEHNPAPAEFLADLISVDYSSSEYVSSQKRVHALYDAFSQSSSVL 361
OY 336 HKTLKNIERMKHLKTLPMVPEKTKDSPGVFSKLVLRV-----TRNLVRNKLAVI 387
Db 362 YATPLSMKEETKNGMRPRKAIVERTDGMWROFFLLKRAMMQASRDGPTNRVARAMSYA 421
OY 388 TRILQNLIMGLFLEFVLNRVSNVLKGAIDRVGLLYOFVGTPTTGMLNAVNLFPVLR 447
Db 422 SA-----VIFGSVFMWKGKSTSIODRMGLLOYAAINTMAAALTKTVGVFPKERA 471
OY 448 VSDQESODGLYOKQMOMLAVALHVLFPFSVATMIFSSVCYWTIGLHPEVARGYFSALL 507
Db 472 IVDREBSKGSYSIGPYLSTKTAETIGAAPPIMRGAVLYPMARLNPILSRGKFCGIYT 531
OY 508 APHLIGEFLTVLGIYONPNIIVNSVALLSIAGYLVSGFLRNIOEMPIPKIISYFTF 567
Db 532 VESFAASAMGLTGVAMVPSTEAAAMAVGSPSLMTVFIYFG-GYVYVNDNTPIIFRWIPRASL 590
OY 568 OKYCEILYVNEFYGLNF 585
Db 591 IRMAPOGICINEFSGLKF 608

```

Search completed: June 11, 2003, 09:13:23
 Job time : 43 secs

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:12:22 ; Search time 25 Seconds

(without alignments)
2688.377 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MGDLSSLPFGSGMGQVNRG.....PALVYIGIVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database: Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB_PEP.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB_PEP.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB_PEP.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB_PEP.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB_PEP.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB_PEP.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3326	100.0	651	US-10-090-455-6	Sequence 6, Appl1
2	3326	100.0	651	US-09-989-981a-6	Sequence 6, Appl1
3	3326	100.0	651	US-09-837-992-3	Sequence 3, Appl1
4	3326	100.0	651	US-09-989-981a-2	Sequence 2, Appl1
5	3326	100.0	651	US-09-837-992-1	Sequence 1, Appl1
6	697	21.0	673	US-09-989-981a-4	Sequence 4, Appl1
7	697	21.0	673	US-10-090-455-7	Sequence 7, Appl1
8	697	21.0	673	US-09-989-981a-8	Sequence 8, Appl1
9	682.5	20.5	655	US-09-981-086-1	Sequence 1, Appl1
10	682.5	20.5	655	US-09-981-353-35	Sequence 35, Appl1
11	680.5	20.5	655	US-10-120-687-61	Sequence 61, Appl1
12	674.5	20.3	655	US-10-090-455-5	Sequence 5, Appl1
13	674.5	20.3	655	US-09-866-866a-10	Sequence 10, Appl1
14	672.5	20.2	655	US-09-866-866a-27	Sequence 27, Appl1
15	660	19.8	657	US-09-866-866a-14	Sequence 14, Appl1
16	602.5	18.1	663	US-10-108-605-245	Sequence 245, App
17	598.5	18.0	674	US-10-090-455-4	Sequence 4, Appl1
18	590.5	17.8	638	US-10-072-621-10	Sequence 10, Appl1
19	585.5	17.6	646	US-10-072-621-9	Sequence 9, Appl1

20	585.5	17.6	646	US-10-090-455-2	Sequence 2, Appl1
21	578.5	17.4	637	US-10-090-455-8	Sequence 8, Appl1
22	578	17.4	604	US-09-745-763-197	Sequence 197, App
23	570.5	17.2	646	US-10-154-452-4	Sequence 4, Appl1
24	567.5	17.1	646	US-10-090-455-13	Sequence 13, Appl1
25	562.5	16.9	646	US-10-154-452-8	Sequence 8, Appl1
26	485	14.6	545	US-10-083-357-1335	Sequence 1335, App
27	403.5	12.1	1564	US-09-801-368-244	Sequence 244, App
28	398	12.0	1501	US-09-801-368-346	Sequence 346, App
29	374	11.2	1511	US-09-801-368-250	Sequence 250, App
30	270.5	8.1	306	US-09-738-626-4554	Sequence 4554, App
31	268	8.1	242	US-09-769-787-83	Sequence 83, Appl1
32	263.5	7.9	345	US-09-815-242-11068	Sequence 11068, A
33	263	7.9	241	US-09-815-242-11194	Sequence 11194, A
34	241.5	7.3	247	US-09-738-626-5108	Sequence 5108, Ap
35	240	7.2	1279	US-10-097-340-2	Sequence 2, Appl1
36	240	7.2	1280	US-09-866-866a-2	Sequence 2, Appl1
37	240	7.2	1280	US-09-866-866a-4	Sequence 4, Appl1
38	240	7.2	1280	US-10-156-239-8	Sequence 8, Appl1
39	239.5	7.2	2436	US-10-199-485-8	Sequence 8, Appl1
40	239.5	7.2	2436	US-09-795-693-8	Sequence 8, Appl1
41	239.5	7.2	1594	US-09-971-121-4	Sequence 4, Appl1
42	238.5	7.2	1642	US-09-971-121-2	Sequence 2, Appl1
43	238.5	7.1	385	US-09-769-787-144	Sequence 144, App
44	236.5	7.1	203	US-10-252-819-17	Sequence 17, Appl1
45	236	7.1	9		

ALIGNMENTS

RESULT 1	US-10-090-455-6	Sequence 6, Application US/10090455
1	MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
2	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
3	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
4	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
5	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
6	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
7	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
8	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
9	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
10	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
11	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
12	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
13	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
14	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
15	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
16	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
17	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
18	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
19	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
20	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
21	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
22	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
23	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
24	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
25	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
26	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
27	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
28	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
29	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
30	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
31	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
32	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
33	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
34	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
35	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
36	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
37	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
38	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
39	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
40	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
41	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
42	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
43	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
44	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60
45	1 MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSGLIHLASTVSHRRPMDITSC	60

QY 301 YMDLTSVDPQSKEREIETSKRVOMIESAKKSAICHKTLKNIERKHLKTLPMVPFKTKD 360
 DB 301 YMDLTSVDPQSKEREIETSKRVOMIESAKKSAICHKTLKNIERKHLKTLPMVPFKTKD 360
 QY 361 SPGVSKLGLVLLRRTRNLVRRKLAIVTRLLONLIMGFLLEFVLRRVSNVTKGAIODRV 420
 DB 361 SPGVSKLGLVLLRRTRNLVRRKLAIVTRLLONLIMGFLLEFVLRRVSNVTKGAIODRV 420
 QY 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
 DB 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
 QY 481 IFSSVCYMTLGLHPEVARGYFSALLAPHLIGFELTVLGLIVONPNIVNSVALLSTA 540
 DB 481 IFSSVCYMTLGLHPEVARGYFSALLAPHLIGFELTVLGLIVONPNIVNSVALLSTA 540
 QY 541 GVLVSGFLRNIOEMPFIKTIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
 DB 541 GVLVSGFLRNIOEMPFIKTIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
 QY 601 AFTOGIOFIEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651
 DB 601 AFTOGIOFIEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651

RESULT 2

US-09-989-981A-6
 Sequence 6, Application US/09989981A
 Publication No. US20030049730A1

GENERAL INFORMATION:

APPLICANT: Hobbs, Helen H.
 APPLICANT: Shan, Bel
 APPLICANT: Barnes, Robert
 APPLICANT: Tian, Hui
 APPLICANT: Tularik Inc.
 APPLICANT: Board of Regents, The University of Texas System
 TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
 FILE REFERENCE: 018781-007320US
 CURRENT APPLICATION NUMBER: US/09/989,981A
 PRIOR FILING DATE: 2002-07-23
 PRIOR APPLICATION NUMBER: US 60/252,235
 PRIOR FILING DATE: 2000-11-20
 PRIOR APPLICATION NUMBER: US 60/253,645
 PRIOR FILING DATE: 2000-11-28
 NUMBER OF SEQ ID NOS: 13
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 6
 LENGTH: 651
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: human ABCG5 (hABCG5)
 US-09-989-981A-6

Query Match 100.0%; Score 3326; DB 9; Length 651;
 Best Local Similarity 100.0%; Pred. No. 1.6e-289;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLQVNRGSSSLEGAAPATAPEPHSIGILHASTVSHRRPMMWDITSC 60
 DB 1 MGDLSLTPGSGMGLQVNRGSSSLEGAAPATAPEPHSIGILHASTVSHRRPMMWDITSC 60
 QY 61 ROOMTRQILKDVSLVYESGOIMCIIIGSSGSKTLLDMSGRIGRAGFTLGEVYVNGRAL 120
 DB 61 ROOMTRQILKDVSLVYESGOIMCIIIGSSGSKTLLDMSGRIGRAGFTLGEVYVNGRAL 120
 QY 121 RROFODCFSVYLOSDTLSSLIVRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
 DB 121 RROFODCFSVYLOSDTLSSLIVRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
 QY 181 ADRLIGNSLGIGISGERRRYSIAOQLDOPKVMLEDEPTGLDCMTANOIVLLVELAR 240
 DB 181 ADRLIGNSLGIGISGERRRYSIAOQLDOPKVMLEDEPTGLDCMTANOIVLLVELAR 240

DB 181 ADRLIGNSLGIGISGERRRYSIAOQLDOPKVMLEDEPTGLDCMTANOIVLLVELAR 240
 QY 241 RRRIVVLIHOPRSELEFOLFDKIALISGELIFCTCPAEMLEDFNDCCGCPCEHSNPFDE 300
 DB 241 RRRIVVLIHOPRSELEFOLFDKIALISGELIFCTCPAEMLEDFNDCCGCPCEHSNPFDE 300
 QY 301 YMDLTSVDPQSKEREIETSKRVOMIESAKKSAICHKTLKNIERKHLKTLPMVPFKTKD 360
 DB 301 YMDLTSVDPQSKEREIETSKRVOMIESAKKSAICHKTLKNIERKHLKTLPMVPFKTKD 360
 QY 361 SPGVSKLGLVLLRRTRNLVRRKLAIVTRLLONLIMGFLLEFVLRRVSNVTKGAIODRV 420
 DB 361 SPGVSKLGLVLLRRTRNLVRRKLAIVTRLLONLIMGFLLEFVLRRVSNVTKGAIODRV 420
 QY 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
 DB 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
 QY 481 IFSSVCYMTLGLHPEVARGYFSALLAPHLIGFELTVLGLIVONPNIVNSVALLSTA 540
 DB 481 IFSSVCYMTLGLHPEVARGYFSALLAPHLIGFELTVLGLIVONPNIVNSVALLSTA 540
 QY 541 GVLVSGFLRNIOEMPFIKTIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
 DB 541 GVLVSGFLRNIOEMPFIKTIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
 QY 601 AFTOGIOFIEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651
 DB 601 AFTOGIOFIEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651

RESULT 3

US-09-837-992-3
 Sequence 3, Application US/09837992
 Patent No. US20020081687A1

GENERAL INFORMATION:

APPLICANT: Tian, Hui
 APPLICANT: Schultz, Joshua
 APPLICANT: Shan, Bel
 APPLICANT: Tularik Inc.
 TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
 FILE REFERENCE: 018781-006020US
 CURRENT APPLICATION NUMBER: US/09/837,992
 PRIOR FILING DATE: 2001-04-18
 PRIOR APPLICATION NUMBER: US 60/198,465
 PRIOR FILING DATE: 2000-04-18
 PRIOR APPLICATION NUMBER: US 60/204,234
 PRIOR FILING DATE: 2000-05-15
 NUMBER OF SEQ ID NOS: 45
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 3
 LENGTH: 651
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
 US-09-837-992-3

Query Match 100.0%; Score 3326; DB 10; Length 651;
 Best Local Similarity 100.0%; Pred. No. 1.6e-289;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLQVNRGSSSLEGAAPATAPEPHSIGILHASTVSHRRPMMWDITSC 60
 DB 1 MGDLSLTPGSGMGLQVNRGSSSLEGAAPATAPEPHSIGILHASTVSHRRPMMWDITSC 60
 QY 61 ROOMTRQILKDVSLVYESGOIMCIIIGSSGSKTLLDMSGRIGRAGFTLGEVYVNGRAL 120
 DB 61 ROOMTRQILKDVSLVYESGOIMCIIIGSSGSKTLLDMSGRIGRAGFTLGEVYVNGRAL 120
 QY 121 RROFODCFSVYLOSDTLSSLIVRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
 DB 121 RROFODCFSVYLOSDTLSSLIVRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180


```

Db      121 REEQFQDCFSYVLQSDTLLSLTIVETLHTLALIRKQNGSFQKKEAVAAELSLRIV 180
Qy      181 ADRLIGNTSLGISTGERRRVSIAAQLQDPRKMLFDEPTTGDCMTANQIVLVLELAR 240
Db      181 ADRLIGNTSLGISTGERRRVSIAAQLQDPRKMLFDEPTTGDCMTANQIVLVLELAR 240
Qy      241 RNRIVVLTIHOPRSLFQDPAKIALISFELICGTPAEMLDFFNDCGYPCEHSNPDE 300
Db      241 RNRIVVLTIHOPRSLFQDPAKIALISFELICGTPAEMLDFFNDCGYPCEHSNPDE 300
Qy      301 YMDLTSVDQSKEREIETSKRVOMESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
Db      301 YMDLTSVDQSKEREIETSKRVOMESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
Qy      361 SPGVFSKLGVLRRVTRNLVNRKLAVITRLLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Db      361 SPGVFSKLGVLRRVTRNLVNRKLAVITRLLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Qy      421 GLLYQFVGATPYTGMLNAAVNFVPLRAVSQDESQDGLYQKQOMLAAVALHVLPPSVATM 480
Db      421 GLLYQFVGATPYTGMLNAAVNFVPLRAVSQDESQDGLYQKQOMLAAVALHVLPPSVATM 480
Qy      481 IFSSVCYWTGLHPEVARFGEYSALLAPHLIGFELTVLGLVQNPNTVNSVVALLSIA 540
Db      481 IFSSVCYWTGLHPEVARFGEYSALLAPHLIGFELTVLGLVQNPNTVNSVVALLSIA 540
Qy      541 GVLVSGFLRNIOEMPIPKIISYFFQKYCSILVYNEFGLNFTGSSNVSATTPMNC 600
Db      541 GVLVSGFLRNIOEMPIPKIISYFFQKYCSILVYNEFGLNFTGSSNVSATTPMNC 600
Qy      601 AFQOGIOFIEKTCGATSRFTANFLIYSFIPALVILGIYVEKIRDLISR 651
Db      601 AFQOGIOFIEKTCGATSRFTANFLIYSFIPALVILGIYVEKIRDLISR 651

```

RESULT 4

US-09-989-981a-2

Sequence 2, Application US/09989981A

Publication No. US20030049730A1

GENERAL INFORMATION:

```

APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-00732005
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2

```

```

Query Match      82.5%; Score 2744.5; DB 9; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
Qy      1 MCDLSLTPGGSMGLQVNRGSSQSLGAPATAPDP-HSLGIILHASVSHRRRPPMDITS 59
Db      1 MGLPFLSPGARGHNRGSSQSLGAPATAPDP-HSLGIILHASVSHRRRPPMDITS 60

```

```

Qy      60 CROQWTRQILKDVSLYVESGOIMCITGSSGCKTLLDAMSGRLGACTFLGEVYNGRA 119
Db      61 COQKMDROILKDVSLYIESGOIMCITGSSGCKTLLDAMSGRLGACTFLGEVYNGRA 120
Qy      120 LREQFQDCFSYVLQSDTLLSLTIVETLHTLALIRKQNGSFQKKEAVAAELSLRIV 179
Db      121 LREQFQDCFSYVLQSDTLLSLTIVETLHTLALIRKQNGSFQKKEAVAAELSLRIV 180
Qy      180 VADRLIGNTSLGISTGERRRVSIAAQLQDPRKMLFDEPTTGDCMTANQIVLVLELAR 239
Db      181 VADRLIGNTSLGISTGERRRVSIAAQLQDPRKMLFDEPTTGDCMTANQIVLVLELAR 240
Qy      240 RNRIVVLTIHOPRSLFQDPAKIALISFELICGTPAEMLDFFNDCGYPCEHSNPDE 299
Db      241 RNRIVVLTIHOPRSLFQDPAKIALISFELICGTPAEMLDFFNDCGYPCEHSNPDE 300
Qy      300 FYMDLTSVDQSKEREIETSKRVOMESAYKKSALCHKTLKNIERKHLKTLPMVPFKTK 359
Db      301 FYMDLTSVDQSKEREIETSKRVOMESAYKKSALCHKTLKNIERKHLKTLPMVPFKTK 360
Qy      360 DSPGVFSKLGVLRRVTRNLVNRKLAVITRLLQNLIMGLFELFVLRRVSNVLKGAIDQDR 419
Db      361 DSPGVFSKLGVLRRVTRNLVNRKLAVITRLLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Qy      420 VGLLYQFVGATPYTGMLNAAVNFVPLRAVSQDESQDGLYQKQOMLAAVALHVLPPSVAT 479
Db      421 VGLLYQFVGATPYTGMLNAAVNFVPLRAVSQDESQDGLYQKQOMLAAVALHVLPPSVAT 480
Qy      480 MIFSSVCYWTGLHPEVARFGEYSALLAPHLIGFELTVLGLVQNPNTVNSVVALLSI 539
Db      481 MIFSSVCYWTGLHPEVARFGEYSALLAPHLIGFELTVLGLVQNPNTVNSVVALLSI 540
Qy      540 AGVLVSGFLRNIOEMPIPKIISYFFQKYCSILVYNEFGLNFTGSSNVSATTPMNC 599
Db      541 AGVLVSGFLRNIOEMPIPKIISYFFQKYCSILVYNEFGLNFTGSSNVSATTPMNC 600
Qy      600 CAFQOGIOFIEKTCGATSRFTANFLIYSFIPALVILGIYVEKIRDLISR 651
Db      601 CAFQOGIOFIEKTCGATSRFTANFLIYSFIPALVILGIYVEKIRDLISR 652

```

RESULT 5

US-09-837-992-1

Sequence 1, Application US/09837992

Patent No. US20020081687A1

GENERAL INFORMATION:

```

APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bel
APPLICANT: Tularik Inc.
TITLE OF INVENTION: SLC12A1 and Methods of Use
FILE REFERENCE: 018781-00602005
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse SLC12A1 and Methods of Use
US-09-837-992-1

```

```

Query Match      82.5%; Score 2744.5; DB 10; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

```


Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapp

Page 5

```

Query Match          21.0%; Score 697; DB 9; Length 673;
Best Local Similarity 28.9%; Pred. No. 9,5e-54;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

OY      8  TRGSGAGLAVNKGSSGLEGAPAT--APEHSHIGLHASTVSYHNR--PMWD-ITSCROOW 64
DB      16  TQDPTSGLODRLEPSSSDNSLYFTTSGQNTLEVRDNLQVLAQVPEFEOLOAFKPMW 75
OY      65  TNGOI-----LKDVSLVESGQIWCITIGSSSGSKTTLIDAMSRLRAGTF--GGEYV 115
DB      76  TSPSCNSCELGIONLSFKVRSGOMALIGSSGCCRASLDVYTR--GHGKIKGQIWT 134
OY      116  NGRALARREGFOCFCESTVYQSLTSLSYREFLHTALLAT--RRCNPGSPFOKVEAVNAE 174
DB      135  NQPPSPOLVRCVAVHRQHNLNLVRETLAEIAOMRPFRTFSQAOROKREVDYIAE 194
OY      175  LSLSHVADRLIGNYSIGISTGERRRVSIADLODDPKVMEDEFETGLDCTANQIYVL 234
DB      195  LRLRQCADTRVGNMNYRGLSGEKRRRVISIGVOLWNPGLITLDEFTSGIDSEFTANLWKT 254
OY      235  LVELARRNRIVVLTTHQPSSELEFQLEDKIALISPEGLFCSTPAEMLDPFNDGCPCEH 294
DB      235  LSRLLKGNKLVYISLHOPRSDIFRFLFDLVLLMTSTPIYLGAQHMVYFAIGPCRY 314
OY      295  SNPFPEYDNLVSVDNOSKERELFENSKRQMIESAATKKAICHKLNIERMKHL----- 348
DB      315  SNPADFVYDLHSIDRRSREOEIAREKQSLAALE-----LEKRDLDDEFLMK 362
OY      349  -----KTLPM-----VPEFKRDPGVESKGLVLRRTNRLVYNKLAIVITRL 390
DB      363  AETKDLDEDTGVESVTFPLDNTCUPSPRK--MGAVOQTTIIRKOISNDFDLPTLLHG 421
OY      391  LQNLINGFLFLFVLARVANSNLKGAIO---DRVGLLYQFVGAPPYTGMLNANLFLPYLR 446
DB      422  AEACIMSTWITGFLYFG-----HGSIQSFMDPTALLFMIGALIPFNVIIDVISKYSER 475
OY      447  AVSDQESODDGLYOKRWQMLAYALAHVLPSPVAVMIFSSVCWTGLHPEVARF----- 499
DB      476  AMLYTELEDGTYTGPYFAKILDELBEPCHAYIITIGPNTYMANLRGLOPFLDLHFLV 535
OY      500  -----GYSALLAPHLIGEFLLVLAGIYQNPNIYNSVALLSTINGVLGSGFL 549
DB      536  WLVEVCCRIMALAAALLPETHMASFES-----NALYNSFYLAG-----GFM 577
OY      550  RNIDMPYFKIITISYTFQKCSILVYNEFYGLNETGGSNNVSVTN 597
DB      578  INLSLMTVPAMISKVSFLRWCFEGELMKIOFSRRTYKMPJLGNLTIAVS 625

RESULT 8
US-09-989-981A-8
: Sequence 8, Application US/0998981A
: Publication No. US20030049730A1
: GENERAL INFORMATION:
:   APPLICANT: Hobbs, Helen H.
:   APPLICANT: Shan, Bei
:   APPLICANT: Barnes, Robert
:   APPLICANT: Tian, Hui
:   APPLICANT: Tularik Inc.
:   APPLICANT: Board of Regents, The University of Texas System
:   TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
:   FILE REFERENCE: 018781-007320US
:   CURRENT APPLICATION NUMBER: US/09/989, 981A
:   PRIOR FILING DATE: 2002-07-23
:   PRIOR APPLICATION NUMBER: US 60/252, 235
:   PRIOR FILING DATE: 2000-11-20
:   PRIOR APPLICATION NUMBER: US 60/253, 645
:   NUMBER OF SEQ ID NOS: 13
:   SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 673

```

```

: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: OTHER INFORMATION: human ABCG8 (hABCG8),
US-09-989-981A-8

```

```

Query Match      21.0%; Score 697; DB 9; Length 673;
Best Local Similarity 28.9%; Pred. No. 5e-54;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

QY      8 TGGGNGGLDVNRGSSSLEGADAT-APEPHSIGILHASYSVSHRV-R-PMD-DITSCROQM 64
           || :||| |-----| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      16 TRDPTDSIGLDRLFESSESDNSLTFFTSQGNPLETRVDNLNQYDLASQWFMQLAGFKMFW 75
           || :||| |-----| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      65 TROI-----LKQVSLYESGOIMCICLGSSGSCKTLLDAMSRLGRAGTF-LGEYVV 115
           | :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      76 TSPSCONCELOIOMLSFERNRSGOMALIISSGCCRASBLDYIVTER-GHGKIKSGOIWI 134
           | :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      116 NGRALRREOFQDCFSFYVLQSDTLISLTIFRETHTALLAI-RIGNPGSPDKVEAVAAE 174
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      135 NCQPSPQLYRCVAHVNOHNLLENLVRETIATAOKRLPRFTSQAKORRRVEDVIAE 194
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      175 LSLSHVADRLIGNYSLGIGSTGERRRVSIAAQLADPKVMLEDEPTTGIDCMANQIYYVL 234
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      195 LRLROCACTRGNMMVRGILSGGERRRVSIGVOLLNPNPILIIDEPSTGLSDFTAHNLVKYT 254
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      235 LVELARRNRIVYLTHOPRSLEFOLEDKIALISFELLFCGTBAMLDPFNDGCPCPEH 294
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      255 LSRLAKGNRLVISLSHPRSSDFLERLEDVLLMTSGPTIYIGAAGHMVOYFTAIGPCPRY 314
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      295 SNPDFEYMLDTSVDTOSKERETENSKRYOMIESAVKRKAICHKTLENERKHL----- 348
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      315 SNPADFVYDLNISDRSRREGELATEREKQSIALAE-----LEKRDLDDELKMK 362
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      349 -----KITLM-----VPFKTKDSPGVFSKIGVLLRRVTNTLVNKLAIVTRL 390
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      363 AEWTDLDDETCVESSVTPFLDNCLPSPTK-MPGAVOOFTTLIRQISINDFRDLPTLLIHG 421
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      391 LQNLIMGLELFEFVLRAVSNVLKGAIQ---DRNGILXQFPFGAPPYIGMLNAVNLFEVLR 446
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      422 AEACIMSMTIFLYFG-----HSIQISFDPTALLPLMGALIPFNVLIDVISKYSER 475
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      447 AVSDQESODGLYORKWMALAYALHVLPEFSVATMIFFSSVCYWTGLHPEVARF----- 499
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      476 AMLEYELEBDGYTTGPYPFAKLIDELPEHCAXIIITYGMPTWLANLARGLQPLHLHLV 535
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      500 -----GFSAALLAPHLICEFLTLYLLGIYQNMNVNSVALLISAGVLVGSGFL 549
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      536 WLWVFCCRIMALAAALLPPFHMASFPS-----NALYNSFYLAG-----GFM 577
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
QY      550 RNIOEMPPIEFKIISYTFQKYCSLIIVNEFYGLNFGCGSSNSVYTN 597
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|
DB      578 INLSIMTVPMWISKVSFLRMCFEGCLKIKQPSRRTYTMPLGNLTIAYS 625
           || :||| :| :||| :| :||| :| :||| :| :||| :| :||| :| :||| :|

RESULT 9
US-09-961-086-1
Sequence 1, Application US/09961086
Publication No. US20030036645A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
APPLICANT: ROSS, Douglas D.
APPLICANT: DOYLE, L. Austin
APPLICANT: ABRUZZO, Lynne
TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
TITLE OF INVENTION: WHICH ENCODES IT
FILE REFERENCE: EPI9376-019
CURRENT APPLICATION NUMBER: US/09/961,086
CURRENT FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: US 60/073,763
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US99/02577
PRIOR FILING DATE: 1999-02-05
```

NUMBER OF SEQ ID NOS: 7
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 655
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-961-086-1

Query Match 20.5%; Score 682.5; DB 9; Length 655;
 Best Local Similarity 29.2%; Pred. No. 1.8e-52;
 Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSLGILHASYSVSHRVRPMDITSCROQMTROIILNDVSLYE 77
 13 SOGNTGFPATASNDKAFTEGAVLSFHNICYRVKLKSGFLPCRKEVEKILLNINGIMK 72
 78 SQGIMCIISSSGSKTLLDAMSGRLGRAGTFIGEYVNGRARRQDPDCSYVQSDT 137
 73 PG-LNALIPPTGGKSSLLDVLAAKRDPSG-LSGDVLLNG-APRPANFCNSGYVQDDV 129
 138 LLSLTVRETLHTALLAIRGNPG-SFOKVEAVAEISLHVADRLIGNYSIGISGTG 196
 130 VNGTLVRENLOPESALRLATMTNHEKNERIRVIOELGDKVADSKYGTOTIRVSGG 189
 197 ERRRVSIQAOLLODPKVMLEDEFTGLDCMTANQIYVLLVELARRNRIVYLTIHQPSREL 256
 190 EKRRTSIGMELITDPSILFLEDEPTGLDSTANAVALLKRMKSGKRTIIFSIHQPRYSI 249
 257 FOLFKAIALISGCELIFCGTPAEMDFNDGCPCEHSPNPFYMDLTSVDTQ---SK 312
 250 FULFSLTLASGRLEHGPADALGYFESAGHCAVANNPAPDFLIDINGSTAVALLNR 309
 313 ERE-----IETSKR-----VOMIESAYKRSATCHKT-----LKNTERMKHLKTLMPVPE 356
 310 EEDFRATEIIEPSKODKPLIEKLAELIYVSSPYKKEKALHOLSGEKKKKTIVVEISY 369
 357 KTKDSPGVSKGLVLRVTRNLVNRKLAIVITRLLONLIMGFL--LFFVLRYRSVNLVG 414
 370 TT-----SFCHQLRWMSKRSFKMLGNPQASIAQIITYVVLGVIGAIYGLKNDST----- 421
 415 ATODRGILYOPVAGAPTYGMLNAVLPVULRVAVSODESDGLYQWOMMLAYAL-HVLP 473
 422 GIONRAGVLE-FLTNOCSSYSVALEFVEKELFHEITISGYRVSSTFLKLSLDLP 480
 474 FSVVATMIFSSVCYWTGLGHPVARGFSALLAPHLIGEFLLVLTGIVONPNIVNSV 533
 481 MMLPSIIFTCIYFIMLGKPKADAFVWMTFLM--WVAYSASSMALAIAGOSVSYVA 537
 534 VALLSIAGY--LVGSGFLNIQEMPIPEKIIISYTFQKCYSEILVYNEFYGLNFTCGSSN 591
 538 TLMATICFVFMMIFSGLLVNLTTIASWLSMLOFYFISIPRYGTALQHNELGONFCPG--- 594
 592 VSVTNPMACAFQGIQIEKTCPG 615
 595 LNATGNPCNTA-----TCTG 610

RESULT 10
 US-09-981-353-35
 Sequence 35, Application US/09981353
 Patent No. US70020160382A1
 GENERAL INFORMATION:
 APPLICANT: Lasek, Amy W.
 TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
 FILE REFERENCE: PA-0038 US
 CURRENT APPLICATION NUMBER: US/09/981,353
 NUMBER OF SEQ ID NOS: 194
 SOFTWARE: PRTL Program
 SEQ ID NO 35
 LENGTH: 655
 TYPE: PRT

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
 US-09-981-353-35

Query Match 20.5%; Score 680.5; DB 9; Length 655;
 Best Local Similarity 29.2%; Pred. No. 2.7e-52;
 Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSLGILHASYSVSHRVRPMDITSCROQMTROIILNDVSLYE 77
 13 SOGNTGFPATASNDKAFTEGAVLSFHNICYRVKLKSGFLPCRKEVEKILLNINGIMK 72
 78 SQGIMCIISSSGSKTLLDAMSGRLGRAGTFIGEYVNGRARRQDPDCSYVQSDT 137
 73 PG-LNALIPPTGGKSSLLDVLAAKRDPSG-LSGDVLLNG-APRPANFCNSGYVQDDV 129
 138 LLSLTVRETLHTALLAIRGNPG-SFOKVEAVAEISLHVADRLIGNYSIGISGTG 196
 130 VNGTLVRENLOPESALRLATMTNHEKNERIRVIOELGDKVADSKYGTOTIRVSGG 189
 197 ERRRVSIQAOLLODPKVMLEDEFTGLDCMTANQIYVLLVELARRNRIVYLTIHQPSREL 256
 190 EKRRTSIGMELITDPSILFLEDEPTGLDSTANAVALLKRMKSGKRTIIFSIHQPRYSI 249
 257 FOLFKAIALISGCELIFCGTPAEMDFNDGCPCEHSPNPFYMDLTSVDTQ---SK 312
 250 FULFSLTLASGRLEHGPADALGYFESAGHCAVANNPAPDFLIDINGSTAVALLNR 309
 313 ERE-----IETSKR-----VOMIESAYKRSATCHKT-----LKNTERMKHLKTLMPVPE 356
 310 EEDFRATEIIEPSKODKPLIEKLAELIYVSSPYKKEKALHOLSGEKKKKTIVVEISY 369
 357 KTKDSPGVSKGLVLRVTRNLVNRKLAIVITRLLONLIMGFL--LFFVLRYRSVNLVG 414
 370 TT-----SFCHQLRWMSKRSFKMLGNPQASIAQIITYVVLGVIGAIYGLKNDST----- 421
 415 ATODRGILYOPVAGAPTYGMLNAVLPVULRVAVSODESDGLYQWOMMLAYAL-HVLP 473
 422 GIONRAGVLE-FLTNOCSSYSVALEFVEKELFHEITISGYRVSSTFLKLSLDLP 480
 474 FSVVATMIFSSVCYWTGLGHPVARGFSALLAPHLIGEFLLVLTGIVONPNIVNSV 533
 481 MMLPSIIFTCIYFIMLGKPKADAFVWMTFLM--WVAYSASSMALAIAGOSVSYVA 537
 534 VALLSIAGY--LVGSGFLNIQEMPIPEKIIISYTFQKCYSEILVYNEFYGLNFTCGSSN 591
 538 TLMATICFVFMMIFSGLLVNLTTIASWLSMLOFYFISIPRYGTALQHNELGONFCPG--- 594
 592 VSVTNPMACAFQGIQIEKTCPG 615
 595 LNATGNPCNTA-----TCTG 610

RESULT 11
 US-10-120-687-61
 Sequence 61, Application US/10120687
 Publication No. US20030082155A1
 GENERAL INFORMATION:
 APPLICANT: Massachusetts General Hospital
 TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating Diabetes Mellitus
 FILE REFERENCE: 3284/12358
 CURRENT APPLICATION NUMBER: US/10/120,687
 PRIOR FILING DATE: 2002-04-11
 PRIOR APPLICATION NUMBER: US60/169082
 PRIOR FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: US 09/963,875
 PRIOR FILING DATE: 2001-09-25
 PRIOR APPLICATION NUMBER: US 60/215109
 PRIOR FILING DATE: 2000-06-28
 PRIOR APPLICATION NUMBER: US 60/238880

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapb

Page 7

;; PRIOR FILING DATE: 2000-10-06
;; PRIOR APPLICATION NUMBER: US 09/731261
;; PRIOR FILING DATE: 2000-12-06
;; NUMBER OF SEQ ID NOS: 61
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 61
;; LENGTH: 655
;; TYPE: prt
;; ORGANISM: Homo sapiens
US-10-120-687-61

Query Match 20.5%; Score 680.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 2,7e-52;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

QY 21 SSSLEGAPATAP---EPHSIGILHASYSVSHRVRPMDITSCROOITROIKDVSLEYE 77
DB 13 SGGNTNGFPATASNDLKAFTEGAVLSPFNICRVKLGFLPCRPEKEILSNINGIMK 72
QY 78 SGQIMCIGSSGSGCTTLLDAMSGRLGAGTFFLGEVYNGRALRREOPDCEFSYVLSPT 137
DB 73 PG-LNALIGPTGGGSSLLDYLAAKDPG--LSGDVLLNG--APRANFKCNGGYVDDV 129
QY 138 LLSLTVRETLHYTALLAIRGNPG-SFOKVEANMAELSHVADRILGNTSLGISTG 196
DB 130 VMGTLTVRENTQFSALALATTTNNHEKERNINRYIELGLDKVADSKVGTQPIFGVSG 189
QY 197 ERRRVSAOILLDOPKVALFDEPTTGLDCAITANOIVLVLELARRNRIYVLIHOPREL 256
DB 190 ERKRTSIGEMELITPILDEPTTGLDSTANANVLLKRMKSGKRTIISIHOPRISI 249
QY 257 FOLDKAIIISFGLIFGCTPAEMIDFFNDCGYPCEPHSNPDEYMDITSVDTQ---SK 312
DB 250 FKLPDSITLLASGRIMFHPAQAELGYFESAGYHCEAVNPADEFLLIINDSTAVALNR 309
QY 313 ERE-----IETSR-----VOMISAYKKAICHKT-----LNIEEMKHLKTLPMYF 356
DB 310 EEDKATEIIPSPKODPLIEKLAETIYNSFYETAELHQLSGEKKKITYFEKISY 369
QY 357 KTKDSPGVSKLGLVLRVTNLRNKLAVTTRLLQNLIMGLFL--LEFVLRVRSNVLK 414
DB 370 TT-----SECHQLRNVKSRSPFNILGNQASIAQIIVYVGLVIGAITFGKLNST---- 421
QY 415 AIDRVGLIYOFVATPTTGLNANVLEPVLRAVSDDESODGLYOKQMMLAYAL-HVLP 473
DB 422 GIORAGVLF-FLTTNOCFSSVSAVELFVEKKLFIEHYISGYRVSSYFLGKLLDLP 480
QY 474 FSVYATMIFSSVCWTGLHPEVARPGYFSALLAPHLIGELTLVLGLIVQNPNIYNSV 533
DB 481 MRMLPSTIFTCIYIFMGLKPKADAFVMMETL---MVAYSASSMALAIAGOSVVA 537
QY 534 VALISAGV--LVSGFLRNIOEMPIPKIISYFTPOKYSIELVYNEFGVLTNCCSSN 591
DB 538 TLMTICFVFMIMISGLVNLITIASLSMLQIYSIRYGTALQHNHEFLGQNCPCG--- 594
QY 592 VSVTNPACAFPTGCIQIEKTCPCG 615
DB 595 LNAATGNPCMYA-----TCTG 610

RESULT 12
US-10-090-455-5
;; Sequence 5, Application US/10090455
;; Publication No. US20030027259A1
;; GENERAL INFORMATION:
;; APPLICANT: Chen, Hongyun
;; TITLE OF INVENTION: NOVEL ABC4 TRANSPORTER AND USES THEREOF
;; FILE REFERENCE: 100103.406
;; CURRENT APPLICATION NUMBER: US/10/090,455
;; PRIOR FILING DATE: 2002-03-01
;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 5
;; LENGTH: 655
;; TYPE: prt
;; ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 20.3%; Score 674.5; DB 9; Length 655;
Best Local Similarity 29.0%; Pred. No. 9,5e-52;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SSSLEGAPATAP---EPHSIGILHASYSVSHRVRPMDITSCROOITROIKDVSLEYE 77
DB 13 SGGNTNGFPATASNDLKAFTEGAVLSPFNICRVKLGFLPCRPEKEILSNINGIMK 72
QY 78 SGQIMCIGSSGSGCTTLLDAMSGRLGAGTFFLGEVYNGRALRREOPDCEFSYVLSPT 137
DB 73 PG-LNALIGPTGGGSSLLDYLAAKDPG--LSGDVLLNG--APRANFKCNGGYVDDV 129
QY 138 LLSLTVRETLHYTALLAIRGNPG-SFOKVEANMAELSHVADRILGNTSLGISTG 196
DB 130 VMGTLTVRENTQFSALALATTTNNHEKERNINRYIELGLDKVADSKVGTQPIFGVSG 189
QY 197 ERRRVSAOILLDOPKVALFDEPTTGLDCAITANOIVLVLELARRNRIYVLIHOPREL 256
DB 190 ERKRTSIGEMELITPILDEPTTGLDSTANANVLLKRMKSGKRTIISIHOPRISI 249
QY 257 FOLDKAIIISFGLIFGCTPAEMIDFFNDCGYPCEPHSNPDEYMDITSVDTQ---SK 312
DB 250 FKLPDSITLLASGRIMFHPAQAELGYFESAGYHCEAVNPADEFLLIINDSTAVALNR 309
QY 313 ERE-----IETSR-----VOMISAYKKAICHKT-----LNIEEMKHLKTLPMYF 356
DB 310 EEDKATEIIPSPKODPLIEKLAETIYNSFYETAELHQLSGEKKKITYFEKISY 369
QY 357 KTKDSPGVSKLGLVLRVTNLRNKLAVTTRLLQNLIMGLFL--LEFVLRVRSNVLK 414
DB 370 TT-----SECHQLRNVKSRSPFNILGNQASIAQIIVYVGLVIGAITFGKLNST---- 421
QY 415 AIDRVGLIYOFVATPTTGLNANVLEPVLRAVSDDESODGLYOKQMMLAYAL-HVLP 473
DB 422 GIORAGVLF-FLTTNOCFSSVSAVELFVEKKLFIEHYISGYRVSSYFLGKLLDLP 480
QY 474 FSVYATMIFSSVCWTGLHPEVARPGYFSALLAPHLIGELTLVLGLIVQNPNIYNSV 533
DB 481 MRMLPSTIFTCIYIFMGLKPKADAFVMMETL---MVAYSASSMALAIAGOSVVA 537
QY 534 VALISAGV--LVSGFLRNIOEMPIPKIISYFTPOKYSIELVYNEFGVLTNCCSSN 591
DB 538 TLMTICFVFMIMISGLVNLITIASLSMLQIYSIRYGTALQHNHEFLGQNCPCG--- 594
QY 592 VSVTNPACAFPTGCIQIEKTCPCG 615
DB 595 LNAATGNPCMYA-----TCTG 610

RESULT 13
US-09-866-866A-10
;; Sequence 10, Application US/09866866A
;; Patent No. US20020102244A1
;; GENERAL INFORMATION:
;; APPLICANT: Sorrentino, Brian
;; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
;; FILE REFERENCE: 1340-1-021CIP2
;; CURRENT APPLICATION NUMBER: US/09/866,866A
;; PRIOR FILING DATE: 2001-08-30
;; PRIOR APPLICATION NUMBER: 09/584,586
;; PRIOR FILING DATE: 2000-05-31
;; PRIOR APPLICATION NUMBER: PCT/US99/11825
;; PRIOR FILING DATE: 1999-05-27
;; PRIOR APPLICATION NUMBER: 60/086,988
;; PRIOR FILING DATE: 1998-05-28
;; NUMBER OF SEQ ID NOS: 27

PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 657
TYPE: PRT
ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 19.8%; Score 660; DB 10; Length 657;
Best Local Similarity 28.0%; Pred No. 1.9e-50;
Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

```
OY 13 MGLQVNGSSSLECAPATAPPEHSLGILHASYSVSHRVRPMWDITSCROQWTRQLIKDY 72
DB 12 MSQRNNNGLPFRNNSRAVRLTLAGDVLSPFHITRYV--KXSGFLV---KTYEKELSDI 66
OY 73 SLXVESGOIMCILGSSGSKTTLIDAMSGRLGAGTGLGEVYVNGRALRREQFODCSYV 132
DB 67 NQIMKPG-LMAILGPTGGKSSLDVLAARKDPKG-LSGDVLNG-APQPAHKKCCSGYV 123
OY 133 LQSDTLISLAVRETLHYTALALAIRGNNGSPQ-----KKVAVNAELSLSHVADRLIGN 187
DB 124 VQDDVVMGLTYRENLQPSALRL-----PTMKNEKNERINTIIEIGLEKRVADSKYGT 179
OY 188 YSLGISTGERRRVSIAQLLODPKVMLEDEPTTGIDCMTANQIVLVLELARBNRIYVL 247
DB 180 QFIRGISGGERKRTSIGMELIDPSILFLEDEPTGLDSTANAVALLIKRMSKQGRITIF 239
OY 248 TIHORSEIFOLFDXIALISGELIFCGTPAPMLDEFNDGCGPCPEHNSNPFDFYMDLTSV 307
DB 240 SIHQRYSIFKLFDSLTLLASGKLVFHPGPAKALEYFASAGYHCPYNNPADFLDYING 299
OY 308 PTOS-----KEREIETSR-----VOMIESAYKKSALCHTKLNIERMKHUKTLP 352
DB 300 DSSAVMLNREDDNEANKTEEPSKGEKPVIENTLSEFYINSALYGETKAEILDQ----- 352
OY 353 MHPFKTSDSPGVFSKLGV-----LRLRVTRNLVRNKLAVITRLQNL 394
DB 353 -----PGAQEKKTSAKEPEYVTSFCHQLRIARSKKNLGNPQASVADLIYTV 403
OY 395 IMGLEFL--LFTVLAVRSNVLKGAIODRVGLYQFVGATPYTGMNAVNLFPVLRASDOE 452
DB 404 ILGLITGAIYFDLKYDA---AGMQNRAGVLE-FLTTNOCFSSVASVELFVEYKELFHE 458
OY 453 SODGIYOKKQMLAYAL-HVLPFSVATMISSVCYWTGLHPEVARFGYSALLAPHL 511
DB 459 YISGIYRVSSYFPGKVMSDLPMRFLPSVIFTCILYFMJLKKTYDAFIMMFTLI---M 515
OY 512 IGEFTLLVLLGIYQNPRIVNSVALLAGV--LVGSGELRNIOEMPIPKIISYFTRK 569
DB 516 VAYTASSMALAIAGOSVSVATILMTIAFVPMLEFSLVNLKRTIGPMLSMLOYESIPR 575
OY 570 YCSEILVNEFEYGLNFTCGSSNVSTYTNPMCAFTOGIOFIKTCPG 615
DB 576 YGFTALDYNEFLGOEFCPG---FNVTDNSTCVNSYAL-----CTG 612
```

Search completed: June 11, 2003, 09:14:46
Job time : 26 secs

Page 1

OM nucleic - nucleic search, using sw model

(without alignments)
9675.628 Million cell updates/sec

Title: US-09-989-981A-5
Perfect score: 3240

Sequence: 1 gtcaggtgagcagcagg,.....aatattcataaacctatgg 2340

Scoring table: IDENTITY_NUC

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubnba/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubnba/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubnba/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubnba/US06_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubnba/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubnba/PCT05_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubnba/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubnba/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubnba/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubnba/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubnba/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubnba/US09C_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubnba/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubnba/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubnba/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubnba/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubnba/US06_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Query No.	Score	Match Length	ID	Description
1	2340	100.0	2340 9	US-09-837-992-4
2	2340	100.0	2340 11	US-09-989-981A-5
3	1395.6	59.6	2258 9	US-09-837-992-2
4	1365.4	58.4	1959 11	US-09-989-981A-1
5	472	20.6	472 9	US-09-837-992-19
6	249	10.2	249 9	US-09-837-992-7
7	214	9.1	214 9	US-09-837-992-14
8	206	8.8	206 9	US-09-837-992-15
9	203.6	8.7	2669 11	US-09-989-981A-7
10	199.2	8.5	2019 11	US-09-989-981A-3
11	186	7.9	186 9	US-09-837-992-17
12	140	6.0	140 9	US-09-837-992-12
13	139.6	6.0	759 13	US-10-027-633-152155
14	139.6	6.0	759 13	US-10-027-633-152156
15	139.6	6.0	759 13	US-10-027-632-152157
16	139	5.9	139 9	US-09-837-992-16

17	137	5.9	137	9	US-09-837-992-9	Sequence 9, Appl1
18	135.4	5.8	472	11	US-09-918-995-30637	Sequence 13037, A
19	130	5.6	130	9	US-09-837-992-13	Sequence 13, Appl1
20	129	5.5	129	9	US-09-837-992-11	Sequence 11, Appl1
21	122	5.2	122	9	US-09-837-992-8	Sequence 8, Appl1
22	114.2	4.9	2930	10	US-09-954-531-591	Sequence 591, Appl1
23	114.2	4.9	2930	14	US-10-171-581-576	Sequence 276, Appl1
24	114.2	4.9	3201	13	US-10-072-621-5	Sequence 5, Appl1
25	113	4.8	113	9	US-09-837-992-18	Sequence 18, Appl1
26	110.8	4.7	2687	13	US-10-154-452-3	Sequence 3, Appl1
27	110.8	4.7	2687	14	US-10-090-455-12	Sequence 12, Appl1
28	107.6	4.6	2687	13	US-10-154-452-7	Sequence 7, Appl1
29	107.2	4.6	1941	14	US-10-090-455-3	Sequence 3, Appl1
30	107.2	4.6	3455	13	US-10-072-621-4	Sequence 4, Appl1
31	107.2	4.6	3455	14	US-10-090-455-1	Sequence 1, Appl1
32	103	4.4	103	9	US-09-837-992-10	Sequence 10, Appl1
33	101.6	4.3	6043	11	US-09-989-981-9	Sequence 9, Appl1
34	98	4.2	2247	10	US-09-866-866A-26	Sequence 26, Appl1
35	98	4.2	2400	13	US-10-108-605-244	Sequence 244, Appl1
36	98	4.2	2418	11	US-09-961-086-2	Sequence 2, Appl1
37	98	4.2	2574	10	US-09-981-333-34	Sequence 34, Appl1
38	98	4.2	2719	14	US-10-120-667-60	Sequence 60, Appl1
39	98	4.2	2719	14	US-09-866-866A-9	Sequence 9, Appl1
40	96.4	4.1	10330	13	US-10-001-189-68	Sequence 68, Appl1
41	90.4	3.9	2025	10	US-09-866-866A-13	Sequence 13, Appl1
42	88.4	3.8	2223	10	US-09-938-842A-2262	Sequence 2262, Appl1
43	83	3.5	3376	14	US-10-037-270-518	Sequence 918, Appl1
44	82.4	3.5	2788	9	US-09-745-763-196	Sequence 196, Appl1
45	72.8	3.1	427	10	US-09-960-352-12839	Sequence 12839, A

ALIGNMENTS

```

RESULT 1
US-09-837-992-4
Sequence 4, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
FILE OF INVENTION: and Methods of Use
FILE REFERENCE: 018781-0060200S
CURRENT APPLICATION NUMBER: US/09/837,992
PRIORITY FILING DATE: 2001-04-18
PRIORITY APPLICATION NUMBER: US 60/198,465
PRIORITY FILING DATE: 2000-04-18
PRIORITY APPLICATION NUMBER: US 60/204,234
PRIORITY FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 2340
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human sitosterolemia gene (SSG)
NAME/KEY: CDS
LOCATION: (107)..(2062)
OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
OTHER INFORMATION: protein
US-09-837-992-4

```

Query Match	100.0%	Score 2340	DB 9	Length 2340
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2340	Conservative 0	Mismatches 0	Indels 0	Gaps 0
1	GTCAAGTGGACGACGACGAGGCGAGTCTCCACACGAGGGTCCCACTGAGACCACTCTGTGGGA	60		
1	GTCAAGTGGACGACGACGAGGCGAGTCTCCACACGAGGGTCCCACTGAGACCACTCTGTGGGA	60		

OY 61 GGGTCGGCCACGAGAAATTTGGCCAGCTTTGGCTGCTGTGGCCATGGGTGACCTCTC 120
DB 61 GGGTCGGCCACGAGAAATTTGGCCAGCTTTGGCTGCTGTGGCCATGGGTGACCTCTC 120
OY 121 ATCTTTAACCCTGGAGGGTTCATGGGTCTCCAGTAAGACAGAGGCTCCAGAGCTTCT 180
DB 121 ATCTTTAACCCTGGAGGGTTCATGGGTCTCCAGTAAGACAGAGGCTCCAGAGCTTCT 180
OY 181 GGAGGGGGCTCTGGCCACGGCCCGGGAGCTTCACAGCTTGGGATCCTCCATGCTCTCTA 240
DB 181 GGAGGGGGCTCTGGCCACGGCCCGGGAGCTTCACAGCTTGGGATCCTCCATGCTCTCTA 240
OY 241 CAGGTCAGCCACCGCTGAGGCCCTGTGGGACATCATCTTGGCCGAGCAGTGGAC 300
DB 241 CAGGTCAGCCACCGCTGAGGCCCTGTGGGACATCATCTTGGCCGAGCAGTGGAC 300
OY 301 CAGGACATCTCTAAGATGTCTCTTGTACGTGGAGGGGAGATCATGTGCATCT 360
DB 301 CAGGACATCTCTAAGATGTCTCTTGTACGTGGAGGGGAGATCATGTGCATCT 360
OY 361 AGGAGCTCAGGCTCGGGGAAACACGCTGCTGGAGCCCATGTCCGGAGGCTGGGGCG 420
DB 361 AGGAGCTCAGGCTCGGGGAAACACGCTGCTGGAGCCCATGTCCGGAGGCTGGGGCG 420
OY 421 CCGGGGACCTTCTGCGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCGGGAGGAGT 480
DB 421 CCGGGGACCTTCTGCGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCGGGAGGAGT 480
OY 481 CCGAGACTGCTTCTCTAGCTCTGACAGACGACCTGTGAGAGAGCTTCACCGTGC 540
DB 481 CCGAGACTGCTTCTCTAGCTCTGACAGACGACCTGTGAGAGAGCTTCACCGTGC 540
OY 541 CGAGAGCTGACCTACACCGCGCTGTGCGCATCCCGCGGGAGTACCGGGCTCTTCCA 600
DB 541 CGAGAGCTGACCTACACCGCGCTGTGCGCATCCCGCGGGAGTACCGGGCTCTTCCA 600
OY 601 GAAGAAGTGAAGGCCGTATGAGCAGAGTGAAGTGAAGCATGTGGCAGACGACTGTAT 660
DB 601 GAAGAAGTGAAGGCCGTATGAGCAGAGTGAAGTGAAGCATGTGGCAGACGACTGTAT 660
OY 661 TGGCAACTACAGTTTGGGGGCAATTTCCACGGGCTGAGCGCGCGGCTCTCATTCGAC 720
DB 661 TGGCAACTACAGTTTGGGGGCAATTTCCACGGGCTGAGCGCGCGGCTCTCATTCGAC 720
OY 721 CCAGCTGCTCCAGAGTCTAAGGTCTAGTGTGTTGATGAGCCACACAGGCTGTGAC 780
DB 721 CCAGCTGCTCCAGAGTCTAAGGTCTAGTGTGTTGATGAGCCACACAGGCTGTGAC 780
OY 781 CATGACTGCTAATCAGATTTGCTCTCTGCTGTAAGTGGCTCGAGAGACGAAATGT 840
DB 781 CATGACTGCTAATCAGATTTGCTCTCTGCTGTAAGTGGCTCGAGAGACGAAATGT 840
OY 841 GGTTCACACATTCACACGCCGCTGTGAGCTTTTCAAGCTTTTGAACAAATTCAT 900
DB 841 GGTTCACACATTCACACGCCGCTGTGAGCTTTTCAAGCTTTTGAACAAATTCAT 900
OY 901 CCTGAGCTTCGGAGAGCTAATTTCTGTGGACGCGACGAGGAAATTCATTTCTTCA 960
DB 901 CCTGAGCTTCGGAGAGCTAATTTCTGTGGACGCGACGAGGAAATTCATTTCTTCA 960
OY 961 TCACTGCGGTTACCTTGTCTGAACATTTCAAACTTTTGAATTCATATGAGCTGAC 1020
DB 961 TCACTGCGGTTACCTTGTCTGAACATTTCAAACTTTTGAATTCATATGAGCTGAC 1020
OY 1021 GTACGTGATACCCCAAGAGGAAAGGAAATAGAACTTCAAGAGAGTCCAAATGAT 1080
DB 1021 GTACGTGATACCCCAAGAGGAAAGGAAATAGAACTTCAAGAGAGTCCAAATGAT 1080
OY 1081 AGAATCTGCTTCAAGAAATTCAGATTTGTATAAATCTTGAAGAAATTTGAAGAT 1140
DB 1081 AGAATCTGCTTCAAGAAATTCAGATTTGTATAAATCTTGAAGAAATTTGAAGAT 1140

OY 1141 GAAACACCTGAAAACTTACCAATGTTCTTTCAAAACCAAGATTTCTCTGAGTTT 1200
DB 1141 GAAACACCTGAAAACTTACCAATGTTCTTTCAAAACCAAGATTTCTCTGAGTTT 1200
OY 1201 CTCTAAACTGGGTCTCTCTGAGAGAGTGAACAAGAACTTGGTGAATAAGCTGGC 1260
DB 1201 CTCTAAACTGGGTCTCTCTGAGAGAGTGAACAAGAACTTGGTGAATAAGCTGGC 1260
OY 1261 AGTGAATACCGCTCTCTCTGAGAACTGATCAGGTTTCTCTCTCTCTCTCTCTCT 1320
DB 1261 AGTGAATACCGCTCTCTCTGAGAACTGATCAGGTTTCTCTCTCTCTCTCTCTCT 1320
OY 1321 GCGGGTCCGAAAGAAATGTCTAAGGTTCTCTATCCAGAGCCGGTGTCTCTTACCA 1380
DB 1321 GCGGGTCCGAAAGAAATGTCTAAGGTTCTCTATCCAGAGCCGGTGTCTCTTACCA 1380
OY 1381 GTTGTGGGCGCCACCCGTACACAGGATGTAAGCGTGTAAATCTGTTCCGGGCT 1440
DB 1381 GTTGTGGGCGCCACCCGTACACAGGATGTAAGCGTGTAAATCTGTTCCGGGCT 1440
OY 1441 GCGAGCTGTACGCGACAGAGAGTGAAGAGCGGCTCTACAGAACTGGCAGATGATCT 1500
DB 1441 GCGAGCTGTACGCGACAGAGAGTGAAGAGCGGCTCTACAGAACTGGCAGATGATCT 1500
OY 1501 GGCCTATGACCTGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
DB 1501 GGCCTATGACCTGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
OY 1561 GTGCTACTGAGACGCTGGGCTTACATCTCTGAGTGTGCGGATTTGAATTTTCTGCTGC 1620
DB 1561 GTGCTACTGAGACGCTGGGCTTACATCTCTGAGTGTGCGGATTTGAATTTTCTGCTGC 1620
OY 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTTGTGCTACTGTGTAATCTCA 1680
DB 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTTGTGCTACTGTGTAATCTCA 1680
OY 1681 AAATCAAAATATGTAAGTGAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
DB 1681 AAATCAAAATATGTAAGTGAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
OY 1741 ATCTGATTCCTCAGAAATATACAGAAATGCGCATCTCTTAAATCATCAGTATTT 1800
DB 1741 ATCTGATTCCTCAGAAATATACAGAAATGCGCATCTCTTAAATCATCAGTATTT 1800
OY 1801 TACATTCAAAAATATGAGAGTGAATCTTGTAGTCAATGATGCTTACGAGACTGAAT 1860
DB 1801 TACATTCAAAAATATGAGAGTGAATCTTGTAGTCAATGATGCTTACGAGACTGAAT 1860
OY 1861 CACTGTGACAGCTCAAAATGTTCTGTGACACTATTCAAATGTGCTTACTCAAG 1920
DB 1861 CACTGTGACAGCTCAAAATGTTCTGTGACACTATTCAAATGTGCTTACTCAAG 1920
OY 1921 AATTCATTCATTTGAGAAACCTGCCAGGTGCAACATCTAGATTCACATGAATCTTCT 1980
DB 1921 AATTCATTCATTTGAGAAACCTGCCAGGTGCAACATCTAGATTCACATGAATCTTCT 1980
OY 1981 GATTTGTATTTATTTATTTACGCTTGTCAATCTAGAAATGTTGTTCAAAATTAAG 2040
DB 1981 GATTTGTATTTATTTATTTACGCTTGTCAATCTAGAAATGTTGTTCAAAATTAAG 2040
OY 2041 GATTCATTCATTTAGAGTGTAGTGAAGGCAATGGCGGGAATATGAATGAATGCTCCG 2100
DB 2041 GATTCATTCATTTAGAGTGTAGTGAAGGCAATGGCGGGAATATGAATGAATGCTCCG 2100
OY 2101 ACTGTGATGACTGCTCTGAAGCTGTGAAGTGAAGTGGCATGATTTTCTTCTGACAG 2160
DB 2101 ACTGTGATGACTGCTCTGAAGCTGTGAAGTGAAGTGGCATGATTTTCTTCTGACAG 2160
OY 2161 GACATCTCAAGTCTTTTAACTTAAGACTCATTTGTGCTCTTGGATCCAAAGAGGCC 2220
DB 2161 GACATCTCAAGTCTTTTAACTTAAGACTCATTTGTGCTCTTGGATCCAAAGAGGCC 2220
OY 2221 TTGAATGCAATGAAGTGGTTATATGTCCTTGTCTTACAACTTGCAGAGGACATGTGT 2280

Db 2221 TTGATTCAAATGGAAGTGGTTTATAGTCCCTGCTTACAACTTGAGGAGCATGTGT 2280
QY 2281 TATTGGAATTTGACTGACGAGCCAGCAAGATGTAAATTAATTATATAAAGCTATAGG 2340
Db 2281 TATTGGAATTTGACTGACGAGCCAGCAAGATGTAAATTAATTATATAAAGCTATAGG 2340

RESULT 2

US-09-989-981A-5
: Sequence 5, Application US/09989981A
: Publication No. US20030049730A1
: GENERAL INFORMATION:
: APPLICANT: Hobbs, Helen H.
: APPLICANT: Shan, Bel
: APPLICANT: Barnes, Robert
: APPLICANT: Tian, Hui
: APPLICANT: Tularik Inc.
: APPLICANT: Board of Regents, The University of Texas System
: TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
: FILE REFERENCE: 018781-007320US
: CURRENT APPLICATION NUMBER: US/09/989, 981A
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: US 60/252,235
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: US 60/253,645
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 2340
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (107)..(2062)
: OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5

Query Match Best Local Similarity 100.0%; Score 2340; DB 11; Length 2340;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCAGTGGAGCAGCAGGAGGAGTCTGCCAGCGGCTCCCACTGAAGCCACTCTGGGA 60
Db 1 GTCAGTGGAGCAGCAGGAGGAGTCTGCCAGCGGCTCCCACTGAAGCCACTCTGGGA 60
QY 61 GGGTCGGCCAGCAAGAAATTTGCCAGCTTGTGCGCATGGTGACCTCTC 120
Db 61 GGGTCGGCCAGCAAGAAATTTGCCAGCTTGTGCGCATGGTGACCTCTC 120
QY 121 ATCTTGAACCCCGGAGGATCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
Db 121 ATCTTGAACCCCGGAGGATCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
QY 121 ATCTTGAACCCCGGAGGATCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
Db 121 ATCTTGAACCCCGGAGGATCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
QY 181 GGAGGGGCTCTCTGACACCGCCCGGAGGCTCAGAGCCGAGGCTCTCATGCTCTTA 240
Db 181 GGAGGGGCTCTCTGACACCGCCCGGAGGCTCAGAGCCGAGGCTCTCATGCTCTTA 240
QY 241 CAGCTCAGCCACCCGCTGAGGCTCTGTGAGCATCATCTTCCGCGCAGAGTGCAC 300
Db 241 CAGCTCAGCCACCCGCTGAGGCTCTGTGAGCATCATCTTCCGCGCAGAGTGCAC 300
QY 301 CAGGAGATCTCTCAAGATGTCTTCTTGTAGTGAAGCGGCGAGATCATGTGATCT 360
Db 301 CAGGAGATCTCTCAAGATGTCTTCTTGTAGTGAAGCGGCGAGATCATGTGATCT 360
QY 361 AGGAAGCTCAGGCTCCGGGAAACAGAGCTGTGAGCGCATGTCCGGAGAGCTGGGGG 420
Db 361 AGGAAGCTCAGGCTCCGGGAAACAGAGCTGTGAGCGCATGTCCGGAGAGCTGGGGG 420
QY 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTAAAGCGGCGGCTGCCCGGAGAGATT 480
Db 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTAAAGCGGCGGCTGCCCGGAGAGATT 480

Db 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTAAAGCGGCGGCTGCCCGGAGAGATT 480
QY 481 CCAGAGCTGCTTCTCTAGTCTCTGACAGCGACACCCCTGTAGAGCTCAACCGTGGC 540
Db 481 CCAGAGCTGCTTCTCTAGTCTCTGACAGCGACACCCCTGTAGAGCTCAACCGTGGC 540
QY 541 CGAGAGCTGCTCTACACCGCGGCTGTGAGGCTCTCCGCGGCAATCCCGCTCTTCCA 600
Db 541 CGAGAGCTGCTCTACACCGCGGCTGTGAGGCTCTCCGCGGCAATCCCGCTCTTCCA 600
QY 601 GAAGAAAGTGAGGCGCTCATGAGAGAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAG 660
Db 601 GAAGAAAGTGAGGCGCTCATGAGAGAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAG 660
QY 661 TGGCACTACAGCTTGGGGGCAATTTCCAGGGGTGAGCGGCGGCTGTGAGGCTGTGAG 720
Db 661 TGGCACTACAGCTTGGGGGCAATTTCCAGGGGTGAGCGGCGGCTGTGAGGCTGTGAG 720
QY 721 CCAGCTGCTCTAGATCTTGAAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAG 780
Db 721 CCAGCTGCTCTAGATCTTGAAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAG 780
QY 781 CATGACTGTAAATCAGATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 781 CATGACTGTAAATCAGATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 841 GGTTCACACATTCACAGCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 841 GGTTCACACATTCACAGCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 CCGAGCTTCTGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 901 CCGAGCTTCTGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 961 TGACTGCGGTACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 961 TGACTGCGGTACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1021 GTCAGTGGATACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1080
Db 1021 GTCAGTGGATACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1080
QY 1081 AGAATCTGCTTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1140
Db 1081 AGAATCTGCTTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1140
QY 1141 GAAACACCTGAAAGCTTCAAGTGTCTTCAAAAGCAAGTCTCTGAGTCTT 1200
Db 1141 GAAACACCTGAAAGCTTCAAGTGTCTTCAAAAGCAAGTCTCTGAGTCTT 1200
QY 1201 CTCTAACTGGGTGTCTCTGAGAGTGTGAGCAAGCAAGTGTGAGCAAGTGTGAGT 1260
Db 1201 CTCTAACTGGGTGTCTCTGAGAGTGTGAGCAAGCAAGTGTGAGCAAGTGTGAGT 1260
QY 1261 AGTATATACGCTCTCTCAAGATCTGATGAGTGTGCTCTCTCTCTCTCTCTCTCT 1320
Db 1261 AGTATATACGCTCTCTCAAGATCTGATGAGTGTGCTCTCTCTCTCTCTCTCTCT 1320
QY 1321 GGGGGTCCGAAGCAATGTCTTAAAGGCTCTTCAAGAGCCGCTGAGTCTCTTACCA 1380
Db 1321 GGGGGTCCGAAGCAATGTCTTAAAGGCTCTTCAAGAGCCGCTGAGTCTCTTACCA 1380
QY 1381 GTTGTGAGGCGCACCCCTGACAGAGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGT 1440
Db 1381 GTTGTGAGGCGCACCCCTGACAGAGTGTGAGAGGCTGTGAGTGTGAGTGTGAGTGT 1440
QY 1441 GCGAGCTGTACAGCAAGCAAGAGAGTGTGAGAGGCTGTGAGAGGCTGTGAGAGTGT 1500
Db 1441 GCGAGCTGTACAGCAAGCAAGAGAGTGTGAGAGGCTGTGAGAGGCTGTGAGAGTGT 1500
QY 1501 GGCCTATGACATGACAGCT 1560
Db 1501 GGCCTATGACATGACAGCT 1560

[illegible]

QY 898 CATCTGAGCTTGGAGAGCTGATTTCTGTGGACGCCAGGAAATGCTGATTTCTT 957
 DB 841 CATCTGAGCTTGGAGAGCTGATTTCTGTGGACGCCAGGAAATGCTGATTTCTT 900
 QY 958 CAATGACTGCGGTACCTTGTCTGTAACATTTCAACCTTTTACCTTCTATGAGACT 1017
 DB 901 CAATGACTGCGGTACCTTGTCTGTAACATTTCAACCTTTTACCTTCTATGAGACT 960
 QY 1018 GAGCTGAGTGTATCCCAAGCAAGGAGGAAATGAAACCTTCAAGAGAGTCCAGAT 1077
 DB 961 GAGCTGAGTGTATCCCAAGCAAGGAGGAAATGAAACCTTCAAGAGAGTCCAGAT 1020
 QY 1078 GATGAAATGCTGCAAGAAATGCAATTTGCAATTTAACTTTGAAGATATGAAAG 1137
 DB 1021 GCTGAAATGCTGCAAGAAATGCAATTTGCAATTTAACTTTGAAGATATGAAAG 1080
 QY 1138 AATGAAACACCTGAAACGTTACCAATGTTCTTCAAAACCAAGATTTCTGAGT 1197
 DB 1081 AGCAGATGATCTGAAACCTTACCAATGTTCTTCAAAACCAAGATTTCTGAGT 1140
 QY 1198 TTTCTCTAACTGCGGTCTCTGAGAGAGTGAAGAACTTTGTTGAATAAGT 1257
 DB 1141 GTTCGGCAAGCTTGGTCTCTGAGAGAGTGAAGAACTTTGTTGAATAAGT 1200
 QY 1258 GGCAGATGATGAGCTCTCTGAGAGAGTGAAGAACTTTGTTGAATAAGT 1317
 DB 1201 GGCAGATGATGAGCTCTCTGAGAGAGTGAAGAACTTTGTTGAATAAGT 1260
 QY 1318 TCTGCGGTGCGGAAGAAATGCTAAAGGCTGATCCAGAGAGGCTGAGTCTCTTA 1377
 DB 1261 TCTGCGGTGCGGAAGAAATGCTAAAGGCTGATCCAGAGAGGCTGAGTCTCTTA 1320
 QY 1378 CCACTTTTGGGCGCAACCCGCTTCAACAGGAGTGAAGCTGATTTGTTCCGT 1437
 DB 1321 TCACTGTTGGGCGCAACCCGCTTCAACAGGAGTGAAGCTGATTTGTTCCGT 1380
 QY 1438 GCTGCGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
 DB 1381 GCTGCGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1498 GCTGCGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
 DB 1441 GCTGCGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1558 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1617
 DB 1501 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1560
 QY 1618 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1677
 DB 1561 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1620
 QY 1678 CCAAAATCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCTGCT 1737
 DB 1621 CCAAAATCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCTGCT 1680
 QY 1738 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1797
 DB 1681 TGTGTGCTACTGAGAGCTGGGCTTACATCTGAGTTGCCGATTTGATTTTCTC 1740
 QY 1798 TTTTCAATTTCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCT 1857
 DB 1741 TTTTCAATTTCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCT 1800
 QY 1858 TTTTCAATTTCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCT 1917
 DB 1801 CTTTCAATTTCCAAATATAGTCAAGAGTGTAGTGTGCTGCTGATTTGGGGGCT 1860
 QY 1918 AGGAATTTCAATTTAGAGAAACCTGCGAGGCAATCTAGATGATCAATGAACTT 1977
 DB 1861 AGGGGTCCAGTTCATGAGAAACCTGCGAGGCAATCTAGATGATCAATGAACTT 1920
 QY 1978 TCTGATTTTGTATTTATTTTCAAGCTTTGCTGATCTTGAAGATTTGTTTCAAAAT 2037

DB 1921 CCTCATCTTATATGAGGTTTATCCAGAGCTCTGCTATCTTACAGATTTTAAAGT 1980
 QY 2038 AAGGATCATCTTATGAGAGTGTGAAACCATGCTGGGAAATGGAAGT 2091
 DB 1981 CAGGAGTACTGATTTAGCAGATGATTTAAGATGACAGGCAAGGAAAGGTTAATG 2034

RESULT 4
 US-09-989-981a-1
 Sequence 1, Application US/0998981A
 : Publication No. US20030049730A1
 : GENERAL INFORMATION:
 : APPLICANT: Hobbs, Helen H.
 : APPLICANT: Shan, Bel
 : APPLICANT: Barnes, Robert
 : APPLICANT: Tlan, Hui
 : APPLICANT: Tularek Inc.
 : APPLICANT: Board of Regents, The University of Texas System
 : TITLE OF INVENTION: ABC5 and ABC8: Compositions and Methods of Use
 : FILE REFERENCE: 018781-007320US
 : CURRENT FILING DATE: 2002-07-23
 : PRIOR APPLICATION NUMBER: US 60/252,235
 : PRIOR FILING DATE: 2000-11-20
 : PRIOR APPLICATION NUMBER: US 60/253,645
 : NUMBER OF SEQ ID NOS: 13
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 1
 : LENGTH: 1959
 : TYPE: DNA
 : ORGANISM: Mus musculus
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (1)..(1959)
 : OTHER INFORMATION: mouse ABC5 (mABC5)
 US-09-989-981a-1

Query Match 58.4%; Score 1365.4; DB 11; Length 1959;
 Best Local Similarity 81.4%; Pred. No. 0;
 Matches 1595; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

QY 107 ATGGGAGCTCTCATCTTTGACCCCGAGGCTCATGAGTCTCAAGTAACAGAGC 166
 DB 1 ATGGGAGCTCTCTCATCTTTGACCCCGAGGCTCATGAGTCTCAAGTAACAGAGC 60
 QY 167 TCCAGAGCTCTCTGAGAGGCTCTGTCACCGCCCGGAGCT---CACAGCTGGC 223
 DB 61 TCTGAGAGCTCTCTGAGAGGCTCTGTCACCGCCCGGAGCT---CACAGCTGGC 120
 QY 224 ATCTCCATGCTCTCTACAGCTGACCCGCTGAGGCTCTGAGGACATCATCT 283
 DB 121 GTCCTGATGTCCTTACAGGCTGACCAACGCTGTGCGGCTTGTGTGAATCAATTA 180
 QY 284 TCCCGGAGCATGACAGGAGATCTCAAAAGATGTCCTTGTAGCTGAGAGCGG 343
 DB 181 TCCCGGAGCATGACAGGAGATCTCAAAAGATGTCCTTGTAGCTGAGAGCGG 240
 QY 344 CAGATCATGTCATCTTGAAGAGCTCAGGCTCCGAGGAAACACAGCTCTGAGCGCATG 403
 DB 241 CAGATCATGTCATCTTGAAGAGCTCAGGCTCCGAGGAAACACAGCTCTGAGCGCATG 300
 QY 404 TCCGGAGGCTGGGGCGGCGGAGCTTCTTGGGGAGGTATGTAGTGAAGCGCGGGG 463
 DB 301 TCCGGAGGCTGGGGCGGCGGAGCTTCTTGGGGAGGTATGTAGTGAAGCGCGGGG 360
 QY 464 CTGGGCGGAGGAGCTTCAAGAGCTGCTTCTCAAGCTCTGACAGAGGAGACCTGCTG 523
 DB 361 CTGGGCGGAGGAGCTTCAAGAGCTGCTTCTCAAGCTCTGACAGAGGAGGCTTCTG 420
 QY 524 AGCAGCTCAGCGGCGGAGAGCTGACATCAACCGGCTGCTGAGGCTCAGCGCGGCG 583

```

Db      421 AGCAGCTCAGTGTGGCGGAGACGTTGCGATACACGCGATGCTGGCCCTCTGCGCAGC 480
OY      584 AATCCGCGCTCTCTCCAGAGAGAGTGAGGCGCGTATGGCAGAGCTGATGACCCAT 643
Db      481 TCCGGGAGCTTCTACACAAAGAGTAGAGGAGTATGACAGAGTGGAGCTGAGCCAC 540
OY      644 GTGGCAGACGAGTATGGGCACTACACCTTGGGGGAGCTTTCCAGGGTAGCGGCGC 703
Db      541 GTGGGAGCAATGATGTGGAGCTATATTTTGGGGAAATTTCCAGTGGCAGGCGGCC 600
OY      704 CGGGTCTCCATGGCAGCCAGCTGCTCCAGATCTTACGTCATGCTGTTTGAAGGCA 763
Db      601 CGAGTTTCCATGGCAGCCAGCTCTTCCAGAGCCCAAGCATGATGCTAATGAGCA 660
OY      764 ACCAGGCGCTGAGCTGATGATGCTATGATGATGCTGCTCTGCTGAGACCTGCT 823
Db      661 ACCAGGAGCTGAGCTGATGATGCTAATGATGATGCTCTGCTGAGCTGAGCT 720
OY      824 CGCAGGAGCCGATGTGGTCTCACCATTACACAGCCCGTCTGAGCTTTTACAGTC 883
Db      721 CGCAGGAGCCGATGTGTATGTATGATGATGATGATGATGATGATGATGATGATG 780
OY      884 TTTGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
Db      781 TTTGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
OY      944 ATGCTGATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1003
Db      841 ATGCTGATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 900
OY      1004 TTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
Db      901 TTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
OY      1064 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1123
Db      961 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
OY      1124 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1183
Db      1021 GAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
OY      1184 GATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1243
Db      1081 GATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
OY      1244 GTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1303
Db      1141 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
OY      1304 CTCTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1363
Db      1201 CTCTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
OY      1364 GTAGTCTCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1423
Db      1261 GTAGTCTCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
OY      1424 AATCTGTTCCGCTGCTGAGCTGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1483
Db      1321 AATCTGTTCCGCTGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1380
OY      1484 AAGTGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1543
Db      1381 AAGTGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
OY      1544 ATGATTTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1603
Db      1441 ATGATTTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
OY      1604 GGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1663
Db      1501 GGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

```

```

OY      1664 CTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1723
Db      1561 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
OY      1724 GCGGGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1783
Db      1621 TCTGGGCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
OY      1784 AAAATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1843
Db      1681 AAAATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
OY      1844 TTTACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1903
Db      1741 TTTACGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800
OY      1904 TGTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1963
Db      1801 TGTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
OY      1964 TTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2023
Db      1861 TTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
OY      2024 GTTGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2082
Db      1921 GTTGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1959

RESULT 5
US-09-837-992-19
; Sequence 19, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultze, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularek Inc.
; TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 13 of hSSG
US-09-837-992-19

Query Match      20.2%; Score 472; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 3.2e-132;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
US-09-837-992-7
Sequence 7, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultze, Joshua
APPLICANT: Shan, Rei

```

? APPLICANT: Tularik Inc.
? TITLE OF INVENTION: Slicosterolemia Susceptibility Gene (SSG): Compositions
? TITLE OF INVENTION: and Methods of use
? FILE REFERENCE: 018781-00602005
? CURRENT APPLICATION NUMBER: US/09/837,992
? CURRENT FILING DATE: 2001-04-18
? PRIOR APPLICATION NUMBER: US 60/198,465
? PRIOR FILING DATE: 2000-04-18
? PRIOR APPLICATION NUMBER: US 60/204,234
? PRIOR FILING DATE: 2000-05-15
? NUMBER OF SEQ ID NOS: 45
? SOFTWARE: Patentln Ver. 2.1
? SEQ ID NO 7
? LENGTH: 249
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: exon 1 of hSSG
? US-09-837-992-7

```

Query Match	10.6%	Score 249;	DB 9;	Length 249;
Best Local Similarity	100.0%	Pred. No. 8.7e-65;		
Matches 249; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	IGTCAGGTGAGCAGGCGAGGCGATCTGCACAGGGCTCCCACTGAAGCCACTCTGGGGA	60
Db		
QY	1 GTCAAGTGGAGCAGGCGACAGTCTGCACAGGGCTCCCACTGAAGCCACTCTGGGGA	60
Db		
QY	61 GGGTCCGGCCACCAAAAATTTGGCCACAGCTTTGCTGGCTTTGGCCATGGGTGACCCCTC	120
Db		
QY	61 GGGTCCGGCCACCAAAAATTTGGCCACAGCTTTGCTGGCTTTGGCCATGGGTGACCCCTC	120
Db		
QY	121 ATCTTTGACCCCGGAGGGCTCATTGGGTCTCCAAATGAACAGAGGCTCCACAGAGCTCCCT	180
Db		
QY	121 ATCTTTGACCCCGGAGGGCTCATTGGGTCTCCAAATGAACAGAGGCTCCACAGAGCTCCCT	180
Db		
QY	181 GGAGGGGGCTCCTTCGCACACGGCCCGGAGGACCTCACAGGCTTGGGATCCTCCATGCTCCCTA	240
Db		
QY	181 GGAGGGGGCTCCTTCGCACACGGCCCGGAGGACCTCACAGGCTTGGGATCCTCCATGCTCCCTA	240
Db		
QY	241 CAGCGTCAG	249
Db		
QY	241 CAGCGTCAG	249
Db		

RESULT 7
 US-09-837-992-14
 : Sequence 14, Application US/09837992
 : Patent No. US20020081687A1
 : GENERAL INFORMATION:
 : APPLICANT: Tian, Hui
 : APPLICANT: Schultz, Joshua
 : APPLICANT: Shan, Bei
 : APPLICANT: Tularik Inc.
 : TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
 : TITLE OF INVENTION: and Methods of Use
 : FILE REFERENCE: 018781-0060200S
 : CURRENT APPLICATION NUMBER: US/09/837,992
 : CURRENT FILING DATE: 2001-04-18
 : PRIOR APPLICATION NUMBER: US 60/198,465
 : PRIOR FILING DATE: 2000-04-18
 : PRIOR APPLICATION NUMBER: US 60/204,234
 : PRIOR FILING DATE: 2000-05-15
 : NUMBER OF SEQ ID NOS: 45
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ ID NO 14
 : LENGTH: 214
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : OTHER INFORMATION: exon 8 of hSSG
 : US-09-837-992-14

	Query Match	9.18;	Score 214;	DB 9;	Length 214;	
	Best Local Similarity	100.0%;	Pred. No. 3.2e-54;			
	Matches 214;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1011 TGCACCTGAGTCAGTCAATGCCAATAACCAGAAGAAATGAACCTCCAGACAG					1076
Db	1 TGAGACTGAGCTAGTGATGATACCCAAGAACGGGAATTGAAAACCTCCAGAGAG					60
QY	1071 TCCAGATGATAGATCTCGCTACACAAGAAATCAGCAATTTGTCAATPAAACCTTTGAAGAATA					1130
Db	61 TCACAGTAGAATGAAATCTGCGCTACACAAGAAATCAGCAATTTGTCAATPAAACCTTTGAAGAATA					120
QY	1131 TTGAAGAATGAACAACACCTGAAAAAGCTTACCAATGTTCTCTTAAAAACCAAGATTCTC					1190
Db	121 TTGAAGAATGAACAACACCTGAAAAAGCTTACCAATGTTCTCTTAAAAACCAAGATTCTC					180
QY	1191 CTGAGATTCTCTAAACCTGGGTTTCTCTCGAG					1224
	181 CTGAGATTCTCTAAACCTGGGTTTCTCTCGAG					214

```

, RESULT 8
, US-09-837-992-15
, Sequence ID, Application US/09837/992
, Patent No. US20020081687A1
, GENERAL INFORMATION:
, APPLICANT: Tian, Hui
, APPLICANT: Schultz, Joshua
, APPLICANT: Shan, Bei
, APPLICANT: Tularik Inc.
, TITLE OF INVENTION: Stlosterolemia Susceptibility Gene (SSG): Compositions
, TITLE OF INVENTION: And Methods of Use
, FILE REFERENCE: 018781-006020US
, CURRENT APPLICATION NUMBER: US/09/837, 992
, CURRENT FILING DATE: 2001-04-18
, PRIOR APPLICATION NUMBER: US 60/198,465
, PRIOR FILING DATE: 2000-04-18
, PRIOR APPLICATION NUMBER: US 60/204,234
, PRIOR FILING DATE: 2000-05-15
, NUMBER OF SEQ ID NOS: 45
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 15
, LENGTH: 206

```


Mon Jul 28 09:43:28 2003

us-09-989-981a-5.rnpb

Page 9

LOCATION: (1) (2019)
OTHER INFORMATION: mouse ABC8 (mABC8)
US-09-989-981a-3

Query Match 8.5%; Score 199.2; DB 11; Length 2019;
Best Local Similarity 54.0%; Pred. No. 4.9e-49;
Matches 430; Conservative 0; Mismatches 363; Indels 3; Gaps 1;

QY 283 TTCCCGGACAGCTGACAGCAGGAGATCTCAAGATCTCTCTAGTGGAGAGCG 342
DB 234 TAGCAGCAGAGATCTCTGAGCTGGGATCCGAATCTAGCTTCAAGTGGAGAGTGG 293
QY 343 GCAGATCATGATCTCTAGAGAGCTGAGCTCCGGGAAACACAGCTCTGAGCGCAT 402
DB 294 ACAGATCTGGGATCATGAGAGCTGAGCTGGGAGAGAGCTCTACTGATCGATGAT 353
QY 403 GTCCGGGAGAGCTGGGAGAGCTGAGCTCTCTGAGAGAGTGTATGAGAGCGCGGC 462
DB 354 CACAGGACAGGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 413
QY 463 GCTGGGAGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 522
DB 414 CAGTACGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 473
QY 523 GAGCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 582
DB 474 GCGCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 533
QY 583 CAATCCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 639
DB 534 CTTCCTCCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 593
QY 640 CCAATGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 699
DB 594 GAGTGGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 653
QY 700 GGGGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 759
DB 654 CCGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 713
QY 760 GCCAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 819
DB 714 ACCCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 773
QY 820 GCGTGGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 879
DB 774 GGGCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 833
QY 880 GCTCTTACAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 939
DB 834 GCTATTTAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 893
QY 940 GGAATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 999
DB 894 GCAATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 953
QY 1000 TGACTTATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1059
DB 954 GGAATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1013
QY 1060 CTCAGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1075
DB 1014 CTTGAGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1029

RESULT 11
US-09-837-992-17
Sequence 17, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei

APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
FILE REFERENCE: 018781-00602005
CURRENT FILING DATE: US/09/837,992
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 17
LENGTH: 186
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: exon 11 of hSSG
US-09-837-992-17

Query Match 7.9%; Score 186; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 9e-46;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1570 GAGCTGGGCTTACATCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1629
DB 1 GAGCTGGGCTTACATCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 60
QY 1630 CCCCAGCTTAAATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1689
DB 61 CCCCAGCTTAAATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 120
QY 1690 TATATCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1749
DB 121 TATATCAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 180
QY 1750 CCTCAG 1755
DB 181 CCTCAG 186

RESULT 12
US-09-837-992-12
Sequence 12, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
FILE REFERENCE: 018781-00602005
CURRENT FILING DATE: US/09/837,992
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 140
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: exon 6 of hSSG
US-09-837-992-12

Query Match 6.0%; Score 140; DB 9; Length 140;
Best Local Similarity 100.0%; Pred. No. 6.6e-32;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 741 AGGTCATGCTGTTGATGAGCCACACAGGCTTGAGTGCATGCTATCATGATTG 800
DB 1 AGGTCATGCTGTTGATGAGCCACACAGGCTTGAGTGCATGCTATCATGATTG 60
OY 801 TCGTCTCTCTGCTGAGTGCAGGACGACGATGCTGCTTCCACATTCACGAGC 860
DB 61 TCGTCTCTCTGCTGAGTGCAGGACGACGATGCTGCTTCCACATTCACGAGC 120
OY 861 CCCGTCTGAGCTTTTCAG 880
DB 121 CCCGTCTGAGCTTTTCAG 140

RESULT 13

US-10-027-632-152155
Sequence 152155, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152155
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152155

Query Match
Best Local Similarity 99.3%; Score 139.6; DB 13; Length 759;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1431 TTCCCGTCTGCGAGCTGTGACGAGGAGAGTACAGAGCGGCTTACCAAGTGGC 1490
DB 42 TTCCCGTCTGCGAGCTGTGACGAGGAGAGTACAGAGCGGCTTACCAAGTGGC 101
OY 1491 AGATGATGCTGGGCTTATGACGAGTCTCTCCCTTACAGCTTGTGCGACATGATT 1550
DB 102 AGATGATGCTGGGCTTATGACGAGTCTCTCCCTTACAGCTTGTGCGACATGATT 161
OY 1551 TCAGCAGTGTGCTACTGG 1570
DB 162 TCAGCAGTGTGCTACTGG 181

RESULT 14

US-10-027-632-152156
Sequence 152156, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152156
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152156

Query Match
Best Local Similarity 99.3%; Score 139.6; DB 13; Length 759;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1431 TTCCCGTCTGCGAGCTGTGACGAGGAGAGTACAGAGCGGCTTACCAAGTGGC 1490
DB 42 TTCCCGTCTGCGAGCTGTGACGAGGAGAGTACAGAGCGGCTTACCAAGTGGC 101
OY 1491 AGATGATGCTGGGCTTATGACGAGTCTCTCCCTTACAGCTTGTGCGACATGATT 1550
DB 102 AGATGATGCTGGGCTTATGACGAGTCTCTCCCTTACAGCTTGTGCGACATGATT 161
OY 1551 TCAGCAGTGTGCTACTGG 1570
DB 162 TCAGCAGTGTGCTACTGG 181

RESULT 15

US-10-027-632-152157
Sequence 152157, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152157
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152157

Query Match
Best Local Similarity 99.3%; Score 139.6; DB 13; Length 759;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY      1431 TTCCGTCGTCGAGCTGTCAGCCAGAGAGAGTCAAGAGGCTCTACCAAGATGGC 1490
      |||||||
Db      42  TTCCCGTCGTGCGAGCTGTCAGGACAGAGAGAGTCAAGAGGCTCTACCAAGATGGC 101
      |||||||
QY      1491 AGATGATGCTGGCCTATGCACTGCAGCTCCTCCCTTCAGCGTTGTTGCCACCATGATTT 1550
      |||||||
Db      102 AGATGATGCTGGCCTATGCACTGCAGCTCCTCCCTTCAGCGTTGTTGCCACCATGATTT 161
      |||||||
QY      1551 TCAGCACTGTGTGCTACTGG 1570
      |||||||
Db      162 TCAGCACTGTGTGCTACTGG 181
      |||||||
```

Search completed: July 27, 2003, 08:56:04
Job time : 501.926 secs

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 17:02:14 ; Search time 566.197 Seconds

(without alignments)
11156.341 Million cell updates/sec

Title: US-09-989-981A-5
Page: 310

Sequence: 1 gtcaggtgagcagcaggg.....aatatcataaacctatgg 2340

Scoring table: IDENTITY_NUC

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 3000000000
```

Post-processing: Minimum Match 0%

Maximum match 1008
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score	Match length DB ID Description

Description

1	2340	100.0	2340	24	AA022009	Human s10steroidla
2	2340	100.0	2340	25	AA048882	Human ABCG5 DNA.
3	23384	109.9	2316	26	ABR51681	Human ABCG5 CDNA s
4	151920	82.1	3240	24	ABR51681	Human encoding human s
5	14052	60.1	3240	24	ABR51685	Mouse ABCG5 CDNA s
6	1395.6	59.6	3256	24	AA022080	Mouse s10steroidla
7	1365.4	58.4	1959	25	ABR51680	Mouse ABCG5 DNA.
8	1363	58.2	2035	24	ABR51686	CDNA encoding rat

9	1335.8	57.1	1915	24	ABK51684	DNA encoding mouse
10	770.2	32.9	1069	24	ABK51687	cDNA encoding hamster
11	472	20.2	4772	24	AAD22022	Human siltosteroide
12	429.4	18.4	432	21	AACT6065	Human OREF ORF162055
13	268	11.5	281	21	AA294755	Human ATP binding
14	249.6	10.7	5460	24	ABK51683	Human ABCG3 upstre
15	249	10.6	249	24	AAD22010	Human siltosteroide
16	214	9.1	214	24	AAD22017	Human siltosteroide
17	206	8.8	206	24	AAD22018	Human siltosteroide
18	203.6	8.7	2669	25	AAD48883	Human ABCG8
19	199.2	8.5	2019	25	AAAD8881	Human siltosteroide
20	199.2	8.5	2564	24	ABN90022	Mouse ABCG8 DNA.
21	186	7.9	186	24	AAD22020	Mouse clone IMX3-6
22	140	6.0	140	24	AAD22015	Human siltosteroide
23	139	5.9	139	24	AAD22019	Human siltosteroide
24	137	5.9	137	24	AAD22012	Human siltosteroide
25	133	5.7	246	24	ABN24693	Human siltosteroide
26	130	5.6	130	24	AAD22016	Human OREF polynuc
27	129	5.5	129	24	AAD22016	Human siltosteroide
28	126.2	5.4	2525	21	AA288625	Human siltosteroide
29	125.4	5.4	3586	24	AB193933	Silkworm Bm white
30	122.2	5.2	371	25	ABT229948	Mouse ischemic co
31	122	5.2	122	24	AAD22011	Breast cancer mark
32	115.4	4.9	133	16	AAWT1044	Human siltosteroide
33	114.8	4.9	2352	23	ABLO5135	Human gene signatu
34	114.2	4.9	2921	25	ABV75074	Drosophila melanog
35	114.2	4.9	2930	21	AA294747	Human Devg22 homol
36	114.2	4.9	2930	24	AB163321	Human ATP binding
37	114.2	4.9	3201	24	ABV74352	Breast cancer rela
38	113	4.8	113	24	ABV22031	Human ABC transpor
39	110.8	4.7	2687	25	AAD66413	Human siltosteroide
40	110.8	4.7	2687	25	ABV77265	Human ABCG4 transp
41	107.6	4.6	2687	25	ABV77267	Nucleotide sequenc
42	107.2	4.6	1941	24	AA144659	Human ABCG4 transp
43	107.2	4.6	3455	24	ABV74351	Human transporter
44	107.2	4.6	3455	24	AAD66412	Human ABC transpor
45	104.4	4.5	226	21	AA101311	Human ABCG4 transp
						Rat liver toxicolo

ALIGNMENTS

RESULT 1
AAD22009

AC AAD22009;

DT 12-FEB-2002 (first entry)
XY

DE Human siostero laemia susceptibility gene (SSG).
XX

sterol-related disorder: hyperlipidaemia; hypercholesterolaemia; hyper-

xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21; ds.

OS Homo sapiens.

FH	Key	Location/Qualifiers
ET	CDS	107 3003

```
/*tag= a
/product= #lines 200
```

PN WO200179272-A2

PD 25-OCT-2001

PF 18-APR-2001; 2001WO-US12758

PR 18-APR-2000; 2000US-198465P

XX

Query Match	100.0%;	Score 2340;	DB 24;	Length 2340;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2340;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	541	CGAAGCGTGCACCTACACCGCGCTGTGGCCATCCGCGCGGCAATCCGCGCTCTTCCA	600
Db	541	CGAGACGCTCACTACACCGCGCTGTGGCCATCCGCGCGGCAATCCGCGCTCTTCCA	600
QY	601	GAGAAGGTGAGAGCGCTCATGGCAGAGCTGAGTCTGAGCCATGTGGGAGACCGATGAT	660
Db	601	GAAAGAGGTGAGAGCGCGTCTATGGCAGAGCTGAGTCTGAGCCATGTGGGAGACCGATGAT	660
QY	661	TGGCAACTACAGCTTGGGGGGGCAATTTCCAGGGGTGAGCGGCGCGGGGTCTCCATGCAAC	720
Db	661	TGGCAACTACAGCTTGGGGGGGCAATTTCCAGGGGTGAGCGGCGCGGGGTCTCCATGCAAC	720
QY	721	CGAGCGTCCCAAGATCCTTAAGTCATGCTGTTTATGAGCCACACAGCGCTGGACTG	780
Db	721	CGAGCGTCCCAAGATCCTTAAGTCATGCTGTTTATGAGCCACACAGCGCTGGACTG	780
QY	781	CATGACTGCTATTCAGATTTGTGCTCTCTGGTGGACACTGCTCCGAGAACCGAATGCT	840
Db	781	CATGACTGCTATTCAGATTTGTGCTCTCTGGTGGACACTGCTCCGAGAACCGAATGCT	840
QY	841	GGTTCTCACCAATTCACACGCGCGGTCTGAGCTTTTTCAGCTCTTTTGACAAATATGGCAT	900
Db	841	GGTTCTCACCAATTCACACGCGCGGTCTGAGCTTTTTCAGCTCTTTTGACAAATATGGCAT	900
QY	901	CCTGAGCTTCGGAAGAGTCGATTTTCTGTGGCACCGCAGCGGAATGCTGATTTCTCAA	960
Db	901	CCTGAGCTTCGGAAGAGTCGATTTTCTGTGGCACCGCAGCGGAATGCTGATTTCTCAA	960
QY	961	TCAGTCGGGTTACCCCTTGCTCGAATTCAAACCTTTTGACTCTTAATGGAGCTGAC	1020
Db	961	TCAGTCGGGTTACCCCTTGCTCGAATTCAAACCTTTTGACTCTTAATGGAGCTGAC	1020
QY	1021	GTCAGTGGATACCCCAAGCAGGAAGGGAATATAGAAACCTTCAGAGAGTCCAGATGAT	1080
Db	1021	GTCAGTGGATACCCCAAGCAGGAAGGGAATATAGAAACCTTCAGAGAGTCCAGATGAT	1080
QY	1081	AGATCTGCTCTCAAGAAATACGCAATTTGTCTATAAACTTTGAAAGAAATTTGAAAGAT	1140
Db	1081	AGAACTCTCTCAAGAAATACGCAATTTGTCTATAAACTTTGAAAGAAATTTGAAAGAT	1140
QY	1141	GAAACACCTGAAGAAAGTTACCAATGTTCTCTTCAAAACCAAAGTCTCCGCGAGTTT	1200
Db	1141	GAAACACCTGAAGAAAGTTACCAATGTTCTCTTCAAAACCAAAGTCTCCGCGAGTTT	1200
QY	1201	CTCTAACTGGGTGTTCTCTGAGAGAGTGAAGAAACTTGTGAGAAATTAACCTGGC	1260
Db	1201	CTCTAACTGGGTGTTCTCTGAGAGAGTGAAGAAACTTGTGAGAAATTAACCTGGC	1260
QY	1261	AGTATTCAGCGCTCTCTCTTCAAGATCTATCATGGGTTTGTTCCTCTTCTTGCTCT	1320
Db	1261	AGTATTCAGCGCTCTCTCTTCAAGATCTATCATGGGTTTGTTCCTCTTCTTGCTCT	1320
QY	1321	GCGGCTCCGAAGCATGTCTAAAGAGGTGCTATCCAGAGACCGGAGTGTCTCTTACCA	1380
Db	1321	GCGGCTCCGAAGCATGTCTAAAGAGGTGCTATCCAGAGACCGGAGTGTCTCTTACCA	1380
QY	1381	GTTTGTGGGCGCACCCCGTACACAGGCATCTGAACGCTGTGATCTGTTTCCCGTCT	1440
Db	1381	GTTTGTGGGCGCACCCCGTACACAGGCATCTGAACGCTGTGATCTGTTTCCCGTCT	1440
QY	1441	GCGAGCTTCACAGCAACGAGAGAGTCAAGAGCGGCTCTACCAAGATGTGGCAGATGATCT	1500
Db	1441	GCGAGCTTCACAGCAACGAGAGAGTCAAGAGCGGCTCTACCAAGATGTGGCAGATGATCT	1500
QY	1501	GCGCTATGCATGCAGCTCTCCCTTCAGCGTTTGTGCCACCAATGATTTTTCACAGATG	1560
Db	1501	GCGCTATGCATGCAGCTCTCCCTTCAGCGTTTGTGCCACCAATGATTTTTCACAGATG	1560
QY	1561	GTGCTACTGACAGCTGGGCTTACATCTCTGAGAGTGTCCCGAATTTGATATTTTCTGTGTC	1620
Db	1561	GTGCTACTGACAGCTGGGCTTACATCTCTGAGAGTGTCCCGAATTTGATATTTTCTGTGTC	1620

OY 1621 TCTGTGGCCCCCAGTAAATGAGTAATTTCTAATCTTGTCTACTGTGTAATGTCGA 1680
 DB 1621 TCTGTGGCCCCCAGTAAATGAGTAATTTCTAATCTTGTCTACTGTGTAATGTCGA 1680
 OY 1681 AATTCGAATATAGTCAACAGTGTAGTGGCTCTCTGTCATTCGCGGGGGGCTGTGG 1740
 DB 1681 AATTCGAATATATCAACAGTGTAGTGGCTCTCTGTCATTCGCGGGGGGCTGTGG 1740
 OY 1741 ATCTGATTCCTCGAAGAACATACAAAGAAATGCCATTCCTTTTAAATCAATCAATATT 1800
 DB 1741 ATCTGATTCCTCGAAGAACATACAAAGAAATGCCATTCCTTTTAAATCAATCAATATT 1800
 OY 1801 TACATTCGCAAAATATGCAAGTGAATGATTTCTGTACAACTAATGATGCTTCACTGAA 1860
 DB 1801 TACATTCGCAAAATATGCAAGTGAATGATTTCTGTACAACTAATGATGCTTCACTGAA 1860
 OY 1861 CACTTGTGGAGCTCAATATGTTCTGTGACAACTAATGATGCTTCACTGAA 1920
 DB 1861 CACTTGTGGAGCTCAATATGTTCTGTGACAACTAATGATGCTTCACTGAA 1920
 OY 1921 AATTCATTCATGAGAAACCTGCGCAGGTGCAACATCTAGATTCAATGAACCTTCT 1980
 DB 1921 AATTCATTCATGAGAAACCTGCGCAGGTGCAACATCTAGATTCAATGAACCTTCT 1980
 OY 1981 GATTTGTATTCATTTATTCACCTCTGTCATCTGATGAGAAATGTTGTTTCAAAATAG 2040
 DB 1981 GATTTGTATTCATTTATTCACCTCTGTCATCTGATGAGAAATGTTGTTTCAAAATAG 2040
 OY 2041 GGATCAATCTATAGAGAGTGTAGTGAAGAACCTGCGGAAATGAGAGAGAGCTGCGG 2100
 DB 2041 GGATCAATCTATAGAGAGTGTAGTGAAGAACCTGCGGAAATGAGAGAGAGCTGCGG 2100
 OY 2101 ACTGTGATGATGCTGCTGTAACCTGTGAATGAGAGTGCATGATTTCTTCTGACAG 2160
 DB 2101 ACTGTGATGATGCTGCTGTAACCTGTGAATGAGAGTGCATGATTTCTTCTGACAG 2160
 OY 2161 GACATCTCAAGTCTTTTAAACATTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 DB 2161 GACATCTCAAGTCTTTTAAACATTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 OY 2221 TTGAATGATGAGAGTGTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 DB 2221 TTGAATGATGAGAGTGTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 OY 2281 TATTTGGAATATGCTGAGTGGAGGAGCCCAAGATTAATTAATCAATCAATGCTGAG 2340
 DB 2281 TATTTGGAATATGCTGAGTGGAGGAGCCCAAGATTAATTAATCAATCAATGCTGAG 2340

RESULT 2
 AAD48882
 ID AAD48882 standard; DNA; 2340 BP.
 AC AAD48882;
 XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Human ABCG5 DNA.
 XX
 KM ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KM sitosterolemia; hyperlipidemia; hypercholesterolemia; gall stone;
 KM HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KM human; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
 KM ABCG5; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT CDS 107..2062
 FT /tag= a
 FT /product= "hABG5 protein"
 XX
 MN MO200281691-A2.

XX 17-OCT-2002.
 PD 20-NOV-2001; 2001WO-US43823.
 PE 20-NOV-2000; 2000US-252235P.
 PR 28-NOV-2000; 2000US-253645P.
 XX
 PA (TUL-) TULARIK INC.
 PA (TEXA) UNIT TEXAS SYSTEM.
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 DR WPI: 2003-058548/05.
 DR P-PSDB; AAE31704.
 PT New ABCG8 polypeptides and nucleic acids, useful for treating
 PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies -
 PS Claim 11, Page 77; 94pp; English.
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is human
 CC ABCG5 DNA.
 SO Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other;
 Query Match 100.0%; Score 2340; DB 25; Length 2340;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GTCAGGTGAGCAGCAGGAGGCTGCGACAGGCTCCCACTGAAGCAGCTGTGGGA 60
 DB 1 GTCAGGTGAGCAGCAGGAGGCTGCGACAGGCTCCCACTGAAGCAGCTGTGGGA 60
 OY 61 GGGTCCGGCCACCAAGAAATTTGGCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 DB 61 GGGTCCGGCCACCAAGAAATTTGGCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 OY 121 ATCTTTGACCCCGGAGGCTCCATGAGGCTCCCAAGTAACAGAGGCTCCAGAGCTCCCT 180
 DB 121 ATCTTTGACCCCGGAGGCTCCATGAGGCTCCCAAGTAACAGAGGCTCCAGAGCTCCCT 180
 OY 181 GAGAGGGGCTCTGTCACCGCCCGGAGGCTGACAGCTGAGGCTGCTGCTGCTGCTGCTG 240
 DB 181 GAGAGGGGCTCTGTCACCGCCCGGAGGCTGACAGCTGAGGCTGCTGCTGCTGCTGCTG 240
 OY 241 GAGGAGGGGCTCTGTCACCGCCCGGAGGCTGACAGCTGAGGCTGCTGCTGCTGCTGCTG 240
 DB 241 GAGGAGGGGCTCTGTCACCGCCCGGAGGCTGACAGCTGAGGCTGCTGCTGCTGCTGCTG 240
 OY 301 CAGGAGATCTCTCAAGATGCTCTGTTAGCTGAGAGGAGGAGGAGGAGTGTGATCTCT 360
 DB 301 CAGGAGATCTCTCAAGATGCTCTGTTAGCTGAGAGGAGGAGGAGGAGTGTGATCTCT 360
 OY 361 AGGAAGCTCAGGCTCCGGGAAACCAAGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 420
 DB 361 AGGAAGCTCAGGCTCCGGGAAACCAAGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 420
 OY 421 CGCGGGAGACTTCTCTGAGGAGGAGTATGTGAAGCGCGGAGGAGGAGGAGGAGGAGGAG 480
 DB 421 CGCGGGAGACTTCTCTGAGGAGGAGTATGTGAAGCGCGGAGGAGGAGGAGGAGGAGGAG 480
 OY 481 CCAGAGCTGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540

Db	481	CCAGAGCTGCTTCTTCCTACGTCCTGACAGACGACACCCTGCTGAGCAGCCTTCAACCGTGC	540
OY	541	CGAGACGCTGCACTACACCGCGCTCTGGCACTCCGCGCGGCAATCCCGGCTCCGTCCA	600
Db	541	CGAGACGCTGCACTACACCGCGCTCTGGCACTCCGCGCGGCAATCCCGGCTCCGTCCA	600
OY	601	GAAAGAGTGGAGGCGCTCATGCGAGAGCTGATCGAACCTATGCGACAGACCGACTGAT	660
Db	601	GAAAGAGTGGAGGCGCTCATGCGAGAGCTGATCGAACCTATGCGACAGACCGACTGAT	660
OY	661	TGGCAACTACAGCTTGGGGGGCACTTCCACGGGTAGAGGGCGCGGGTCTCCATGCAAC	720
Db	661	TGGCAACTACAGCTTGGGGGGCACTTCCACGGGTAGAGGGCGCGGGTCTCCATGCAAC	720
OY	721	CCAGTGTCTCAGAGTCTTAAGGTCATGCTGTTATGAGGCCAACACAGCGCTGCACTG	780
Db	721	CCAGTGTCTCAGAGTCTTAAGGTCATGCTGTTATGAGGCCAACACAGCGCTGCACTG	780
OY	781	CATGACTGCTAATCAGATTGTGTCCTCTGGTGAACGTGGTCGACGAAACGAATGT	840
Db	781	CATGACTGCTAATCAGATTGTGTCCTCTGGTGAACGTGGTCGACGAAACGAATGT	840
OY	841	GGTTCCTACCACTTACACGACCCCGCTTCGAGCTTTTTCAGCTCTTTGACAAAATGGCAT	900
Db	841	GGTTCCTACCACTTACACGACCCCGCTTCGAGCTTTTTCAGCTCTTTGACAAAATGGCAT	900
OY	901	CCTGAGCTCGAGAGCTGATTTTCTGAGACGCGACGCGGAAACGCTGATTTCTTCAA	960
Db	901	CCTGAGCTCGAGAGCTGATTTTCTGAGACGCGACGCGGAAACGCTGATTTCTTCAA	960
OY	961	TGACTGGGTTACCCCTGTCTCTGAAACATTCAAACCTTTTGACTCTTATGAGACCTGAC	1020
Db	961	TGACTGGGTTACCCCTGTCTCTGAAACATTCAAACCTTTTGACTCTTATGAGACCTGAC	1020
OY	1021	GTGCTGGATTCACCAAGCAAGAAACGGGAAATAGAAACCTCCACAGAGCTCCAGATGAT	1080
Db	1021	GTGCTGGATTCACCAAGCAAGAAACGGGAAATAGAAACCTCCACAGAGCTCCAGATGAT	1080
OY	1081	AGAATCTGCTCACAGAAATCAGCAATTTGTGATTAACCTTGAAGATATTTGAAGAT	1140
Db	1081	AGAATCTGCTCACAGAAATCAGCAATTTGTGATTAACCTTGAAGATATTTGAAGAT	1140
OY	1141	GAAACACCTGAAAACGTTACCAATGTTCTTTCAAAAACCAAGAATTCCTCGAGTTT	1200
Db	1141	GAAACACCTGAAAACGTTACCAATGTTCTTTCAAAAACCAAGAATTCCTCGAGTTT	1200
OY	1201	CTCTAAACCTGGGTTCTTCCTGAGAGAGGTACACAAGAACTTGGTGAAGATAAGCTGCG	1260
Db	1201	CTCTAAACCTGGGTTCTTCCTGAGAGAGGTACACAAGAACTTGGTGAAGATAAGCTGCG	1260
OY	1261	AGTGAATACGCGTCTCCTTCAAGATCTGATGATGGTTTGTTCCTCTTCTTCTGCTCT	1320
Db	1261	AGTGAATACGCGTCTCCTTCAAGATCTGATGATGGTTTGTTCCTCTTCTTCTGCTCT	1320
OY	1321	GCGGGTCCGAAGCAATGTCTAAAGGGTCTATCCAGAGACGCGTAGGTCTCTTTACCA	1380
Db	1321	GCGGGTCCGAAGCAATGTCTAAAGGGTCTATCCAGAGACGCGTAGGTCTCTTTACCA	1380
OY	1381	GTTTGTGGGGGCAACCCCGTACACAGCAATGCTAAAGCTGTGAATCTGTTCCCGTCT	1440
Db	1381	GTTTGTGGGGGCAACCCCGTACACAGCAATGCTAAAGCTGTGAATCTGTTCCCGTCT	1440
OY	1441	GCGAGCTGTACGCAACAGAGAGTCAAGACGGCTCTACACAGAAAGGGCAGATGATGCT	1500
Db	1441	GCGAGCTGTACGCAACAGAGAGTCAAGACGGCTCTACACAGAAAGGGCAGATGATGCT	1500
OY	1501	GGGCTATGCACTGACGTCTCCCTTCAACGGTGTGTCACACACTGATTTTTCACAGCTG	1560
Db	1501	GGGCTATGCACTGACGTCTCCCTTCAACGGTGTGTCACACACTGATTTTTCACAGCTG	1560
OY	1561	GTGCTATGCAAGCTGGGCTTACATCCTGAGAGTTGCCGATTTGATATTTTCTGCTGC	1620
Db	1561	GTGCTATGCAAGCTGGGCTTACATCCTGAGAGTTGCCGATTTGATATTTTCTGCTGC	1620

QY	1621	TCCTGTGGCCCCCACCCTAATTTGGTGAATTTCTAACTCTGTGCTACTTGGTATGCTCA	1680
Db	1621	TCCTGTGGCCCCCACCCTAATTTGGTGAATTTCTAACTCTGTGCTACTTGGTATGCTCA	1680
QY	1681	AAATCCAAATATAGTCAACAGTGTGTGGCTGTGTGTCATATGGGGGGGTGCTTTGGG	1740
Db	1681	AAATCCAAATATAGTCAACAGTGTGTGGCTGTGTGTCATATGGGGGGGTGCTTTGGG	1740
QY	1741	ATCTGGATTCCTCAGAAAACATATCAAGAAATGCCCCATTCCTTTTAAATCATCAGTTATTT	1800
Db	1741	ATCTGGATTCCTCAGAAAACATATCAAGAAATGCCCCATTCCTTTTAAATCATCAGTTATTT	1800
QY	1801	TTCATTTCCAAAATATTCGAGTGAATTCCTGTAGTCAATGAGTCTACGGAGCTGAATTT	1860
Db	1801	TTCATTTCCAAAATATTCGAGTGAATTCCTGTAGTCAATGAGTCTACGGAGCTGAATTT	1860
QY	1861	CAGCTGTGGCAGCTCCAAATGTTTCTGTGTACACATATCATATGTGTGCTTCACCTCAAGG	1920
Db	1861	CAGCTGTGGCAGCTCCAAATGTTTCTGTGTACACATATCATATGTGTGCTTCACCTCAAGG	1920
QY	1921	AATTCATATTCATGAGAAAACCTGCCCCAGGTGCAACATCTAGATTACAAATGAATCTTCT	1980
Db	1921	AATTCATATTCATGAGAAAACCTGCCCCAGGTGCAACATCTAGATTACAAATGAATCTTCT	1980
QY	1981	GATTTTGTATTCATTTATTTCCAGCTCTTGTCATCCTAGAGATGTTGTTTCAAAATTAAG	2040
Db	1981	GATTTTGTATTCATTTATTTCCAGCTCTTGTCATCCTAGAGATGTTGTTTCAAAATTAAG	2040
QY	2041	GGATCATCTCATTTAGCAGGTAGTGAACCCATGCTGGGAAAATGGAATGTAACCTCCCG	2100
Db	2041	GGATCATCTCATTTAGCAGGTAGTGAACCCATGCTGGGAAAATGGAATGTAACCTCCCG	2100
QY	2101	ACTGTGCATGACTGCTCTGCAACGTCGTGAATAGAGAGTCCCATGTTATTTCTTTTGACAG	2160
Db	2101	ACTGTGCATGACTGCTCTGCAACGTCGTGAATAGAGAGTCCCATGTTATTTCTTTTGACAG	2160
QY	2161	GACATCTCTAAAGTCTTTTACCATTTAAGACTCCATTTGTGCTCTTGATCCAAACAGGACC	2220
Db	2161	GACATCTCTAAAGTCTTTTACCATTTAAGACTCCATTTGTGCTCTTGATCCAAACAGGACC	2220
QY	2221	TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTTACAACTTTGACAGGACATGTGGT	2280
Db	2221	TTGAATGCAATGGAAGTGGTTTATAGTCCCTTGCTTACAACTTTGACAGGACATGTGGT	2280
QY	2281	TATTTGGAAATTTGACTGAGCCGAGCCCAAGAAATGTAATTAATTCATTAACCTATAGGG	2340
Db	2281	TATTTGGAAATTTGACTGAGCCGAGCCCAAGAAATGTAATTAATTCATTAACCTATAGGG	2340

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 XX WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 XX sitosterolemia, arteriosclerosis and heart diseases
 PS
 XX Example 3; Page 37-38; 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of human ABCG5 gene located
 CC on chromosome 2p21.
 XX
 XX Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 other;
 SQ
 Query Match 99.9%; Score 2338.4; DB 24; Length 2516;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

455 CGCGGGAGACCTTCCTGGGGGAGGTATATGTAACGGCCGGCGCTCGCCGGGAGCAGTT 514
 481 CCAGAGCTGCTTCTCTACGTCTGTCAGAGCGACACCCCTGTCAGACCTTCACCTGCG 540
 515 CCAGAGCTGCTTCTCTACGTCTGTCAGAGCGACACCCCTGTCAGACCTTCACCTGCG 574
 541 CGAGAGCTGCTTCTCTACGTCTGTCAGAGCGACACCCCTGTCAGACCTTCACCTGCG 600
 575 CGAGAGCTGCTTCTCTACGTCTGTCAGAGCGACACCCCTGTCAGACCTTCACCTGCG 634
 601 GAAGAAGCTGAGGCGCTCATGAGAGAGTGTGAGCATGTGTCACAGCAGAT 660
 635 GAAGAAGCTGAGGCGCTCATGAGAGAGTGTGAGCATGTGTCACAGCAGAT 694
 661 TGCCAACTACAGCTTGGGGGCAATTTCCACGGGTGACGGCGCGGCTTCATCCGACG 720
 695 TGCCAACTACAGCTTGGGGGCAATTTCCACGGGTGACGGCGCGGCTTCATCCGACG 754
 721 CCAGCTGCTCCAGAGATCTTAAGTATGATGCTGTTGATGACCAACACAGGCTGACCTG 780
 755 CCAGCTGCTCCAGAGATCTTAAGTATGATGCTGTTGATGACCAACACAGGCTGACCTG 814
 781 CATGACTGTATATCAGATTTGTGCTCTGTCAGTGTGGAACCTGGTCGAGGAACCGAATGT 840
 815 CATGACTGTATATCAGATTTGTGCTCTGTCAGTGTGGAACCTGGTCGAGGAACCGAATGT 874
 841 GGTTCCTACCATTCACAGAGCCCGTCTGAGCTTTTACAGCTTTTGAACAAATATGCGAT 900
 875 GGTTCCTACCATTCACAGAGCCCGTCTGAGCTTTTACAGCTTTTGAACAAATATGCGAT 934
 901 CCGAGCTTCCAGAGATCTTAAGTATGATGCTGTTGATGACCAACACAGGCTGACCTG 960
 935 CCGAGCTTCCAGAGATCTTAAGTATGATGCTGTTGATGACCAACACAGGCTGACCTG 994
 961 TGACTGCGGTATACCTTGTCTGTCATTAACCTTTGATCTTATATGACCTGAC 1020
 995 TGACTGCGGTATACCTTGTCTGTCATTAACCTTTGATCTTATATGACCTGAC 1054
 1021 GTCAGTGTATACCAAGCAAGAGAGGGAATAGAACTCCCAAGAGCTCAGATGAT 1080
 1055 GTCAGTGTATACCAAGCAAGAGAGGGAATAGAACTCCCAAGAGCTCAGATGAT 1114
 1081 AGAATGTGCTTACAAAGAAATCAGCAATTTGTCTAATAAATTTGAAAGAAAT 1140
 1115 AGAATGTGCTTACAAAGAAATCAGCAATTTGTCTAATAAATTTGAAAGAAAT 1174
 1141 GAACACAGCTGAAGAACTTACCAATGCTTCTTCAAAACCAAGAAATTTCTCTGAGATTTT 1200
 1175 GAACACAGCTGAAGAACTTACCAATGCTTCTTCAAAACCAAGAAATTTCTCTGAGATTTT 1234
 1201 CTTTAAACCTGGGTGTTCTCTGAGAGAGTGAACAAAGCTTGTGAGAAATTAAGCTGGC 1260
 1235 CTTTAAACCTGGGTGTTCTCTGAGAGAGTGAACAAAGCTTGTGAGAAATTAAGCTGGC 1294
 1261 AGTGAATACCGCTCTCTTACAGATGTATCATGAGGTTGTTCTCTTCTCTCTCTCT 1320
 1295 AGTGAATACCGCTCTCTTACAGATGTATCATGAGGTTGTTCTCTTCTCTCTCTCTCT 1354
 1321 GCGGGTCCGAAGCAATGTGCTTAAAGGCTCTATCCAGAGACCGCGTATAGTCTCTTTACCA 1380
 1355 GCGGGTCCGAAGCAATGTGCTTAAAGGCTCTATCCAGAGACCGCGTATAGTCTCTTTACCA 1414
 1381 GTTTGTGGGCGCACCCCTGTCACAGAGCTGCTGAAGAGCTGTGAATCTGTTCCCGTGT 1440
 1415 GTTTGTGGGCGCACCCCTGTCACAGAGCTGCTGAAGAGCTGTGAATCTGTTCCCGTGT 1474
 1441 GCGAGCTGTACAGCAACAGAGAGTGTGAGAGGCTTACAGAAAGTGTGACAGATGCT 1500
 1475 GCGAGCTGTACAGCAACAGAGAGTGTGAGAGGCTTACAGAAAGTGTGACAGATGCT 1534
 1501 GCGCTATGACATGACAGCTCTCCCTTACAGCTTGTGACAGATGATTTTACAGATGT 1560
 1535 GCGCTATGACATGACAGCTCTCCCTTACAGCTTGTGACAGATGATTTTACAGATGT 1594

QY	1561	GTGCTACTGAGCGCTGGGCTTACATCCTGAGGTGCCGATTTGGATTTTTCGTGCTGC	1620
Db	1595	GTGCTACTGAGCGCTGGGCTTACATCCTGAGGTGCCGATTTGGATTTTTCGTGCTGC	1654
QY	1621	TCTGTGGCCCCCACCCTTAATTGGTGAATTTCTACTCTTGTCTACTGTGTATGCTCCA	1680
Db	1655	TCTGTGGCCCCCACCCTTAATTGGTGAATTTCTACTCTTGTCTACTGTGTATGCTCCA	1714
QY	1681	AAATCCAAATTAATGATCAACAGTGTAGTGGCTCTGTCGCAATTCGGGGGGGCTGTGTGG	1740
Db	1715	AAATCCAAATTAATGATCAACAGTGTAGTGGCTCTGTCGCAATTCGGGGGGGCTGTGTGG	1774
QY	1741	ATCTGGAATTCCTCGAAGAACATACAGAAATGCCATTCCTTTTAAATTCATCACTTAATTT	1800
Db	1775	ATCTGGAATTCCTCGAAGAACATACAGAAATGCCATTCCTTTTAAATTCATCACTTAATTT	1834
QY	1801	TACATTCGCAAAAATATATGCAAGTAGATTCCTTTGTATGTCAATGAGTGTCTACGAGCTGAATTT	1860
Db	1835	TACATTCGCAAAAATATATGCAAGTAGATTCCTTTGTATGTCAATGAGTGTCTACGAGCTGAATTT	1894
QY	1861	CACATTGTGGAGCTGCATATGTTTCTGTGACAACTATACCAATGTGTCCTTCACTCAAGG	1920
Db	1895	CACATTGTGGAGCTGCATATGTTTCTGTGACAACTATACCAATGTGTCCTTCACTCAAGG	1954
QY	1921	AATTCAAATTCATTGAGAAAACTGCGCAGGTGCAACATCTAGATTCCAAATGAACCTTCT	1980
Db	1955	AATTCAAATTCATTGAGAAAACTGCGCAGGTGCAACATCTAGATTCCAAATGAACCTTCT	2014
QY	1981	GATTTTGTATTCATTTTATTCACAGCTCTGTTCATCTCTAGGAATAGTGTTTTCAAAATTAAG	2040
Db	2015	GATTTTGTATTCATTTTATTCACAGCTCTGTTCATCTCTAGGAATAGTGTTTTCAAAATTAAG	2074
QY	2041	GGATCATCTGATTTAGCAGTAGTGTAGTAAAGCCATGGCTGGGAAAAATGGAAGTGAAGCTCCG	2100
Db	2075	GGATCATCTGATTTAGCAGTAGTGTAGTAAAGCCATGGCTGGGAAAAATGGAAGTGAAGCTCCG	2134
QY	2101	ACGTGTGATGATACGTCTGTGAACGTCGTCTAATTAAGAGTGGCATGTATTTCTTCTTGACAG	2160
Db	2135	ACGTGTGATGATACGTCTGTGAACGTCGTCTAATTAAGAGTGGCATGTATTTCTTCTTGACAG	2194
QY	2161	GACATCTCAAGCTTTTAAACCATTAAGATCCACATTTTGTGCTCTGTGGATCCAGAGCGCC	2220
Db	2195	GACATCTCAAGCTTTTAAACCATTAAGATCCACATTTTGTGCTCTGTGGATCCAGAGCGCC	2254
QY	2221	TTCGATGCAATGGAAGTGGTTTATTAAGTCCCTTGCCTCTTACACACTTGACAGGAGACATGTGT	2280
Db	2255	TTCGATGCAATGGAAGTGGTTTATTAAGTCCCTTGCCTCTTACACACTTGACAGGAGACATGTGT	2314
QY	2281	TATTTGGAAATTTGAGCTGAGCGGAGCCAGCAAGATGTATAATTAATTCATATAACCTATGGG	2340
Db	2315	TATTTGGAAATTTGAGCTGAGCGGAGCCAGCAAGATGTATAATTAATTCATATAACCTATGGG	2374

RESULT 4	ID	ABK51681	standard; DNA; 1920 BP.	Location/Qualifiers
XX	AC	ABK51681;		1..1920
XX	DT	30-JUL-2002	(first entry)	/tag- a
XX	DE	DNA encoding human ABCG5 protein.		
XX	KM	Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; chromosome 2p21; ds.		
XX	OS	Homo sapiens.		
XX	FT	CDS		
XX	FT			

FT	/product= "Human ABCG5 protein"
FT	/transl_except- (pos: 4..9, aa: GDLSLTPGGSMGL)
XX	/note= "This sequence contains 13 exons"
XX	
PD	WO200227016-A2.
XX	
PD	04-APR-2002.
XX	
PF	25-SEP-2001; 2001WO-US29859.
XX	
PR	25-SEP-2000; 2000US-235268P.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
PA	(PATE/) PATEL S B.
XX	(DEAN/) DEAN M.
PI	Patel SB, Dean M;
DR	WPT: 2002-416483/44.
XX	P-PSDB; AAN98984.
XX	
PT	Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT	nucleic acid encoding the polypeptide, useful for treating
XX	sitosterolemia, arteriosclerosis and heart diseases
PS	Claim 38; Page 36-37; 66pp; English.
XX	
CC	The present invention relates to a new mammalian ATP-binding cassette
CC	gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC	predisposition for developing sitosterolemia, arteriosclerosis or heart
CC	disease. The molecules of the invention are also useful for identifying
CC	a compound which alters ABCG5 activity level comprising contacting a cell
CC	culture or mammal which have ABCG5 polypeptide with a compound and
CC	measuring ABCG5 biological activity in the cell culture or in mammal,
CC	where an increase or decrease in ABCG5 biological activity compared to
CC	ABCG5 biological activity in a control cell culture or mammal not
CC	contacted with the compound, identifies a compound that increases or
CC	decreases ABCG5 activity respectively. The cell culture or mammal
CC	comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC	ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC	polypeptide in a cell culture or mammal is also compared with that of a
CC	second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC	Stimulation of ABCG5 activity is useful for treating or preventing
CC	hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC	disease. The method of the invention is useful for increasing cholesterol
CC	excretion and/or decreasing cholesterol adsorption. The present nucleic
CC	acid sequence represents the human ABCG5 gene located on chromosome 2p21.
CC	This sequence encodes the human ABCG5 protein of the invention.
XX	
SQ	Sequence 1920 BP: 440 A; 503 C; 486 G; 491 T; 0 other;
	Query Match 82.1%; Score 1920; DB 24; Length 1920;
	Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0
	Matches 1920; Conservative 0; Mismatches 0
OY	143 ATGGGTCTCCAAGTAACAGAGAGGCTCCAGAGTCTCCTGAGAGGGGCTCTGCACCAGCC 202
Db	1 ATGGGTCTCCCAAGTAACAGAGAGGCTCCAGAGTCTCCTGAGAGGGGCTCTGCACCAGCC 60
OY	203 CCGGAGGCTCAACACCTGGGGATCTTCATGCTCTTAGACGCTAAGCCACCGCGTGAGG 262
Db	61 .CCGGAGGCTCAACACCTGGGGATCTTCATGCTCTTAGACGCTAAGCCACCGCGTGAGG 120
OY	263 CCCGTGTGGGACATCACATCTTTGCCGGGAGAGTGGAGACAGAGATCTCAAAAGATATGC 322
Db	121 CCCGTGTGGGACATCACATCTTTGCCGGGAGAGTGGAGACAGAGATCTCAAAAGATATGC 180
OY	323 TCCTTTGACGTGAGAGCGGGGAGATCATATGTCATTCTTAGGAAGCTCAGGCTCCGGGAAA 382
Db	181 TCCTTTGACGTGAGAGCGGGGAGATCATATGTCATTCTTAGGAAGCTCAGGCTCCGGGAAA 240
OY	383 ACCACGGTGTGAGAGCGGCATATCTCCGGAGAGGTGGGGGCGCGGGGACTTCTCTGGGGAG 442

Db	241	ACACAGCTGTGGACGGCATTGTCCGGGAGGCTTGGGCGCGCGGGGACCTTTCCTGGGGAG	300
QY	443	GTTATATGTGAACGGCGGGCGCTGGGCGCGGGAGCACTTCCAGACAGCTTCTTCTTCAAGTC	502
Db	301	GTGTATGTGAACGGCGGGCGGCTGGCGCGGGAGCACTTCCAGACAGCTTCTTCTTCAAGTC	360
QY	503	CTGCAGAGCAGACCCCGTGTGACACCTTACCGGTGCGGGAGACGCTGCACCTACACCGCG	562
Db	361	CTGCAGAGCAGACCCCGTGTGACACCTTACCGGTGCGGGAGACGCTGCACCTACACCGCG	420
QY	563	CTGCTGGCCATCCCGCGGGCAATTCGCGGCTCTTCCAGAAAGAAAGGTGGAGGCGGCTAAG	622
Db	421	CTGCTGGCCATCCCGCGGGCAATTCGCGGCTCTTCCAGAAAGAAAGGTGGAGGCGGCTAAG	480
QY	623	GCAGACCTGAGTGTGACCACTGTGGCAGACCGCACTAATTGGCAACTACAGCTTGGGGGCG	682
Db	481	GCAGACCTGAGTGTGACCACTGTGGCAGACCGCACTAATTGGCAACTACAGCTTGGGGGCG	540
QY	683	AATTCCAGGGGTAGGGGGCGGGGCTCATGTGCAGGCCAGCTGTCTCAGAGATTCCTAAG	742
Db	541	AATTCCAGGGGTAGGGGGCGGGGCTCATGTGCAGGCCAGCTGTCTCAGAGATTCCTAAG	600
QY	743	GTTCATGCTGTTTGATGAGCCAACACAGAGCCTGTGATGATGACGTCTAATGAGATTTGC	802
Db	601	GTTCATGCTGTTTGATGAGCCAACACAGAGCCTGTGATGATGACGTCTAATGAGATTTGC	660
QY	803	GTCTCTCTGTGGAACTGGCTGCGAGGAACCGAATTTGTGTTCTCACCATTCCACAGCCC	862
Db	661	GTCTCTCTGTGGAACTGGCTGCGAGGAACCGAATTTGTGTTCTCACCATTCCACAGCCC	720
QY	863	CGTTCGACCTTTTTCAGCTCTTTTGACAAATTTGCCATCTGAGCTTGGAGAGCTGATT	922
Db	721	CGTTCGACCTTTTTCAGCTCTTTTGACAAATTTGCCATCTGAGCTTGGAGAGCTGATT	780
QY	923	TTCTGTGAGACCCAGCGGAAATGCTGATTTCTTCAATGAATGGGGTTAACCTGTCTCT	982
Db	781	TTCTGTGAGACCCAGCGGAAATGCTGATTTCTTCAATGAATGGGGTTAACCTGTCTCT	840
QY	983	GAACTTCAAAACCTTTTGACTTCTATATGACCTGACCTCAGTGGATTACCCAAGCAAG	1042
Db	841	GAACTTCAAAACCTTTTGACTTCTATATGACCTGACCTCAGTGGATTACCCAAGCAAG	900
QY	1043	GAAAGGAAATGAAACCTCCCAAGAGATCCAGATGATGAATCTGCTACAGAATACTA	1102
Db	901	GAAAGGAAATGAAACCTCCCAAGAGATCCAGATGATGAATCTGCTACAGAATACTA	960
QY	1103	GCAATTTGCAATAAACTTGAAGAATTTGAADAATGAAGAAACACTGAAAAAGCTTACCA	1162
Db	961	GCAATTTGCAATAAACTTGAAGAATTTGAADAATGAAGAAACACTGAAAAAGCTTACCA	1020
QY	1163	ATGTTCTCTTTCAAAAACCAAGATTTCTCTGAGATTTTCTAACTGGGTGTTCTCTG	1222
Db	1021	ATGTTCTCTTTCAAAAACCAAGATTTCTCTGAGATTTTCTAACTGGGTGTTCTCTG	1080
QY	1223	AGGAGAGTGACAAAGAACTTGGTGAATAATAGCTGCGAGTGAATTACGCTCTCTTACG	1282
Db	1081	AGGAGAGTGACAAAGAACTTGGTGAATAATAGCTGCGAGTGAATTACGCTCTCTTACG	1140
QY	1283	AATCGATCATGAGGTTTGTCTCTCTTCTTCTCGTTGCGGGGTCCGAAGAAATGTGCTA	1342
Db	1141	AATCGATCATGAGGTTTGTCTCTCTTCTTCTCGTTGCGGGGTCCGAAGAAATGTGCTA	1200
QY	1343	AAGGTTGCTATCCAGAGACCGGTAGGTCCTTATACAGTTTGTGGGCGCACCCCGTAC	1402
Db	1201	AAGGTTGCTATCCAGAGACCGGTAGGTCCTTATACAGTTTGTGGGCGCACCCCGTAC	1260
QY	1403	ACAGGCAATGCTGAACGCTGTGAATCTGTTTCCCGTGTGCGAGCTGTTCAGAGCAAGAG	1462
Db	1261	ACAGGCAATGCTGAACGCTGTGAATCTGTTTCCCGTGTGCGAGCTGTTCAGAGCAAGAG	1320
QY	1463	AGTCAGAGCGGCTCTTACCAAGTGGAGAGATGATCGGGCTAATGCACTGACAGTCTCTC	1522
Db	1321	AGTCAGAGCGGCTCTTACCAAGTGGAGAGATGATCGGGCTAATGCACTGACAGTCTCTC	1380

QY	1523	CCCTTCAGGGTGTGTTGGCCACCATTGATTTTCAGCACTGTGTCTACTGGACGCTGGCTTA	1582
Db	1381	CCCTTCAGGGTGTGTTGGCCACCATTGATTTTCAGCACTGTGTCTACTGGACGCTGGCTTA	1440
QY	1583	CATCCGTGAGGTGCCCCGATTTGGATATATTTTCTGTGTCTCTTGGGCCCCCACTTATTT	1642
Db	1441	CATCCGTGAGGTGCCCCGATTTGGATATATTTTCTGTGTCTCTTGGGCCCCCACTTATTT	1500
QY	1643	GGTGAATTTCTTACTCTGTGTGCTACTTGTGTGATTCGTCCAAAATCCAAATATATGCAACAGT	1702
Db	1501	GGTGAATTTCTTACTCTGTGTGCTACTTGTGTGATTCGTCCAAAATCCAAATATATGCAACAGT	1560
QY	1703	GTAGAGGGCTGTGCTGTCCATTCGCGGGGGGTGCTGTGTGATTCGTGATTCCTCAGAAACATA	1762
Db	1561	GTAGAGGGCTGTGCTGTCCATTCGCGGGGGGTGCTGTGTGATTCGTGATTCCTCAGAAACATA	1620
QY	1763	CAGAAATAGCCCATCTCTTTTAAATCATTCAGATTATTTTACATCCAAAATATTTGGAGT	1822
Db	1621	CAGAAATAGCCCATCTCTCTTTTAAATCATTCAGATTATTTTACATTCCTCAGAAATATTTGAGT	1680
QY	1823	GAGATTCTTGTAGTCATATGAGTTCTACGAGCATGATTTCACTTGTGCGAGCTCAAAATGTT	1882
Db	1681	GAGATTCTTGTAGTCATATGAGTTCTACGAGCATGATTTCACTTGTGCGAGCTCAAAATGTT	1740
QY	1883	TCTGTGACACATTAATCCAAATGTGTGCTCTACTCAGAGAAATTCATTCATTTGAGAAACC	1942
Db	1741	TCTGTGACACATTAATCCAAATGTGTGCTCTACTCAGAGAAATTCATTCATTTGAGAAACC	1800
QY	1943	TGCCAGGTGCAACATCTAGATTTCACAATGAACCTTTCGATTTGTATTCATTTATTCGA	2002
Db	1801	TGCCAGGTGCAACATCTAGATTTCACAATGAACCTTTCGATTTGTATTCATTTATTCGA	1860
QY	2003	GCTCTTGTCACTCAGGATATAGTTGTTTCAAAAATTAAGGATCATCTCATTTAGCAGGTAG	2062
Db	1861	GCTCTTGTCACTCAGGATATAGTTGTTTCAAAAATTAAGGATCATCTCATTTAGCAGGTAG	1920

XX Example 3; Page 45; 66pp; English.

PS The present invention relates to a new mammalian ATP-binding cassette
XX gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of
XX the invention.

SO Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 other;

Query Match 60.2%; Score 1409.2; DB 24; Length 2354;
Best Local Similarity 80.4%; Pred. No 0;
Matches 1664; Conservative 0; Mismatches 403; Indels 3; Gaps 1;

QY 25 CTGCCAGGCGCTCCCAACTGAAAGCCACTGTGGGAGGTCGCGGCCACAGAAATTTGC 84
DB 57 CTCCTCATTTGCTCTCTAGTTAAAGCTCCCTGGAGCGGACAGGCCATGAAATTTAC 116
QY 85 CCAGCTTTGCTGCTGTGGCCATGGGTGACCTCTATCTTTGACCCCGGAGGTCAT 144
DB 117 TTCATTTTCTCTCTCTAGCTAGGTGGTGGAGTCCCTTCTAGTCCAGAGGAGCAG 176
QY 145 GGGTCTCAAGTAAAGAGAGGCTCCAGAGCTCCCTGGAGGGGCTCTCCAGCCGCC 204
DB 177 AAGGCGCTCACTCAAGAGAGGCTCTGAGCTCCCTGGAGCAAGTTCCGTCCAGGCGAC 236
QY 205 GGAAGCCT--CACAGCCTGGGATCTCTCATGCTCTCTACAGCGTACGCCAGCTGAG 261
DB 237 AGAGGCTCGGACAGCTAGGTGTCTGCAATGTGTCTCTACAGCGTACAGCAGTGTGCG 296
QY 262 GCCCTGTGGGAAATCATATCTTCCGCGAGCAGTGGAGCAGCAGATCTCTAAAGATGT 321
DB 297 GCTTGTGGAAATCAATCAATCATGAGCAGCAAGAGTGGAGACAGCAATCTCTCAAGATGT 356
QY 322 CTCCTGTAGTGGAGAGGCGGAGATCATGTGCTATCTTGAAGAGTCAAGGCTCCGCGGA 381
DB 357 CTCCTGTATCATGAGAGAGTGGAGATTTATGTCTATCTTGAAGAGTCAAGGCTCCGCGGA 416
QY 382 AACCAAGCTGTGGAGCAGATGTCTCGGAGGCTGGGCGCGGAGCCTTCTGGGGG 441
DB 417 GACCAAGCTGTGGAGCAGATGTCTCGGAGGCTGGGCGCGGAGCCTTCTGGGGG 476
QY 442 GGTGTATGTGAAGCGCGCGCGCTGTGCGCGGAGCAGTTCAGAGATGCTTCTCTAAGT 501
DB 477 GGTGTGTGAAGGTGTGGAGTGTGGAGGAGCAGTTCAGAGATGCTTCTCTAAGT 536
QY 502 CTGCAAGAGCAGACCTGTGTGAGAGCTCACGCGGAGAGAGCTGTGACAGACCGC 561
DB 537 CTGCAAGAGAGAGCTTTTCTGTAGAGCTCTCACTGTGCGGAGAGCTGTGAGAGACAGC 596
QY 562 GCTGTGGCCATCCGCGCGGAGCAATCCGAGCTCTTTCAGAGAGAGTGGAGGCTCTCAT 621
DB 597 GATGCTGGCCCTGTGCGGAGCTCCGCGGAGCTTCTCAAGAGAGAGTGGAGGAGTCTCAT 656
QY 622 GCGAGAGCTGAGTGTGAGCAGTGTGAGCAGCAGATGATTTGGCACTACAGCTTGGGGG 681

DB 657 GACAGAGCTGAGCTGAGCCAGCTGGCGGAGCAAAATGATTTGAGATTAATTTTGGGGG 716
QY 682 CATTCAGGAGGAGAGGCGCGGCTCTCATTCGAGCCAGCTGCTCAGATCTTA 741
DB 717 AATTTCCAGTGGAGGCGCGGAGTCTTCTATTCGAGCCAGCTCTTATGAGACCCAA 776
QY 742 GGTCAATCTTTGATGAGCAACACAGAGCTGAGTGCATGATCTTAATGATGT 801
DB 777 GGTCAATCTTTGATGAGCAACACAGAGCTGAGTGCATGATCTTAATGATGT 836
QY 802 GTTCTCTGCTGGAGAGCTGGCTGAGAGACCAATTTGTTCTCCATTCACAGCC 861
DB 837 CTTCTCTTGGTGGAGCTGCTGCGAGGAGCCAAATTTGATTTGATTCACATTCACAGCC 896
QY 862 CCGTTCAGCTTTTTCAGCTCTTTGACAAATTTGCAATCTTCTGAGAGCTGAT 921
DB 897 TCGCTCTGAGCTTTTCAACACTTGCACAAATTTGCAATCTTCTGAGAGCTGAT 956
QY 922 TTTCTGTGGAGCCAGGAGAAATGCTGATTTCTTCAATGATGCGGTTACCTGTGC 981
DB 957 GTTCTGTGGAGCCAGAGAGAGTGTGCTGCTTCTCAATGATGCTTGTACCTGTGC 1016
QY 982 TGAACATTCAAACCTTTTGTACTTCTATATGACCTGAGCTGAGTATCCCAAGCA 1041
DB 1017 TGAACATTCCAATCCCTTTGATTTTACATGAGCTTGCATGAGTGGAGCCCAAGAG 1076
QY 1042 GGAAGCGGAATGAATGAATCCCAAGAGAGCTCCAGATGATGATCTGCTCAAGAAATC 1101
DB 1077 AGAGCGGAGAAATGAATGAATCCCAAGAGAGCTCCAGATGATGATCTGCTCAAGAAATC 1136
QY 1102 AGCAATTTGCTATTAATCTTTGAAGATATTTGAAGAAATGAATGAATGAATGAAT 1161
DB 1137 TGAATCTATTCACAAATTTCTGAGAGATTTGAAGAGAGCTGATTAATGAATGAAT 1196
QY 1162 AATGTTCTTTTCAAAACCAAGATTTCTCTGAGATTTTCTTAACTGGTGTCTCTCT 1221
DB 1197 CACGTTCTTTTCAAAACCAAGATTTCTCTGAGATTTTCTTAACTGGTGTCTCTCT 1256
QY 1222 GAGGAGTGAAGAGAAATCTGTTGAGAAATGAAGAGTATTAATGAGTATTAATGAGT 1281
DB 1257 GAGGAGTGAAGAGAAATCTGTTGAGAAATGAAGAGTATTAATGAGTATTAATGAGT 1316
QY 1282 GATCTGATCAATGAGTGTGTTCTCTCTCTTCTGTTGAGGAGTCCGAGCAATGAGT 1341
DB 1317 GATCTGATCAATGAGTGTGTTCTCTCTCTTCTGTTGAGGAGTCCGAGCAATGAGT 1376
QY 1342 AAAGGTGCTATTCAGAGACCGGCTAGTGTCTCTTACAGATTTGTGGGCGCACCCGTA 1401
DB 1377 AAAGGTGCTATTCAGAGACCGGCTAGTGTCTCTTACAGATTTGTGGGCGCACCCGTA 1436
QY 1402 CACAGGATGCTGAAGAGCTGTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1461
DB 1437 CACAGGATGCTGAAGAGCTGTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1496
QY 1462 GAGTCAAGAGAGCTCTTACAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1521
DB 1497 GAGTCAAGAGAGCTCTTACAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1556
QY 1522 CCCCTTCAAGCTTTGTTCCACAGATATTTTACAGAGTGTGCTACTGAGAGCTGGGCT 1581
DB 1557 CCCCTTCAAGCTTTGTTCCACAGATATTTTACAGAGTGTGCTACTGAGAGCTGGGCT 1616
QY 1582 ACATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1641
DB 1617 GATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1676
QY 1642 TGTGTAATTTCAACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1701
DB 1677 TGTGTAATTTCAACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1736
QY 1702 TGTGTAATTTCAACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1761
DB 1737 TGTGTAATTTCAACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796

QY 1762 ACAGAAATGCCATTCCTTTTAAATCATCATGTTATTTACATTCACAAAATATTCAG 1821
 DB 1797 ACAGAAATGCCATTCCTTTTAAATCATCATGTTATTTACATTCACAAAATATTCAG 1856
 QY 1822 TGAGATTCCTGTAGTCAATAGTTCACGACGATTCATTCATTCGACGCTCAATATG 1881
 DB 1857 TGAGATTCCTGTAGTCAATAGTTCACGACGATTCATTCATTCGACGCTCAATATG 1916
 QY 1882 TTCTGTACAACTAATCATCATGTTATTTACATTCACAAAATATTCAG 1941
 DB 1917 CTCTATGCTTAATCACCACGATTCGACGATTCACGACGATTCACGACGATTCACGAC 1976
 QY 1942 CTGCCCAGGTGACAACTAATCATCATGTTATTTACATTCACAAAATATTCAG 2001
 DB 1977 CTGCCCAGGTGACAACTAATCATCATGTTATTTACATTCACAAAATATTCAG 2036
 QY 2002 AGCTCTGTATCTCTAGGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 2061
 DB 2037 AGCTCTGTATCTCTAGGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 2096
 QY 2062 GTGAAGGATGCTGCTGGAAATGGAATG 2091
 DB 2097 GTTAAGATGACAGGACGAAAGGTTATG 2126

RESULT 6
 AAD22008 standard; DNM: 2258 BP.

AC AAD22008;
 DT 12-FEB-2002 (first entry)

DE Mouse sitosterolemia susceptibility gene (SSG).

KM Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
 KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KM xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

OS Mus sp.

FT Key Location/Qualifiers
 FT CDS 47..2005
 FT /tag= a
 FT /product= "Mouse SSG protein"

MO200179272-A2.

25-OCT-2001.

18-APR-2001; 2001WO-US12758.

18-APR-2000; 2000US-198465P.

15-MAY-2000; 2000US-204234P.

(TULAR)-TULARIK INC.

Tian H, Schultz J, Shan B.

WPI; 2002-017598/02.

P-PSDB; ABE13289.

Novel sitosterolemia susceptibility gene polypeptide and
 polynucleotide, useful for screening a compound that increases the
 level of expression or activity of SSG polypeptide for treating
 sterol-related disorder

Claim 8; Fig 7; 105pp; English.

The invention relates to an isolated sitosterolemia susceptibility gene
 (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)

CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic hemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensic and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17.
 XX

SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 other;

Query Match 59.6%; Score 1395.6; DB 24; Length 2258;
 Best Local Similarity 80.7%; Pred. No. 0;
 Matches 1642; Conservative 0; Mismatches 389; Indels 3; Gaps 1;

QY 61 GGGTCCGGCCACAGAAATTTGCCAGCTTTCCTGCTGTTGGCCATGGTACCTTC 120
 DB 1 GGGACAGGCCACTAGAAATTCATTCATTCCTTCCTGCTGTTGGCCATGGTACCTTC 60
 QY 121 ATCTTGACCCCGGAGGGTCCATGGGTCTCCAGATTAACAGAGCTCCAGACCTCCT 180
 DB 61 CTTTCTGATTCAGAGGAGCCAGAGGGCTTCATCAATCAACAGAGGCTCTCTGACCTCC 120
 QY 181 GGAGGGGCTCTGCTCCACCGCCCGGAGCT---CAGACGCTGGGATCTCCATGCTCC 227
 DB 121 GGAGCAAGTTCGGGTACCGGACAGAGGCTCGGACACTTATGATGCTGATGCTG 180
 QY 238 CTACAGCTGACAGCCCGGCTGAGGCTCTGTTGGATCATCATCTTGGCGGACGATG 297
 DB 181 CTACAGCTGACAGCCCGGCTGAGGCTCTGTTGGATCATCATCTTGGCGGACGATG 240
 QY 298 GACGAGGAGATTCCTTAAGATGCTCTTGTACGAGAGGCGGACATCATCTGATG 357
 DB 241 GGACAGGCAATTCCTTAAGATGCTCTTGTACGAGAGGCGGACATCATCTGATG 300
 QY 358 CCTAGGAGCTCAGGCTCCGGGAAACAGCGCTGAGAGGCTCCGGAGGCTGGG 417
 DB 301 CTAGGAGCTCAGGCTCCGGGAAACAGCGCTGAGAGGCTCCGGAGGCTGGG 360
 QY 418 GCGCGGGGACCTTCCTGAGGAGGTATGTAACGCGCGGCTCGCGGAGAGA 477
 DB 361 GCGCACTGGGACCTTCGAAAGGAGGTGTTGTGAATGCTCGAGCTCGCAGGAGACA 420
 QY 478 GTTCCAGAGCTGCTTCTCTACGCTCTGAGAGCGACACCTCTGAGACGCTCACCT 537
 DB 421 GTTCCAGAGCTGCTTCTCTACGCTCTGAGAGCGACACCTCTGAGACGCTCACCT 480
 QY 538 GCGGAGAGCTGACACTACACCGGCTGTGGCCATTCGCGCGGCAATCCCGGCTCTT 597
 DB 481 GCGGAGAGCTGACACTACACCGGCTGTGGCCATTCGCGCGGCAATCCCGGCTCTT 540
 QY 598 CCAGAAAGAGTGAAGCCGCTCATGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAG 657
 DB 541 CACAAAGAGTGAAGCCGCTCATGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAG 600
 QY 658 GATTGGCACTACAGCTTGGGGGCAATTTCCAGGAGTGAAGCGGCGCGGCTTCATG 717
 DB 601 GATTGGCACTACAGCTTGGGGGCAATTTCCAGGAGTGAAGCGGCGCGGCTTCATG 660
 QY 718 AGCCGAGCTGCTCCAGAGCTTAAGGTATGTTATGAGCCCAACAGGCTGGA 777
 DB 661 AGCCGAGCTGCTCCAGAGCTTAAGGTATGTTATGAGCCCAACAGGCTGGA 720
 QY 778 CTGATGAGCTGATTAAGATTCGCTCTGCTGAGAGCTGAGAGCTGAGAGCTGAG 837
 DB 721 CTGATGAGCTGATTAAGATTCGCTCTGCTGAGAGCTGAGAGCTGAGAGCTGAG 780

```

OY 838 TGTGTTCTCACCATTTCACACGCCCCGTTCTGAGCTTTTTCAGCTTTTGACAAATTCG 897
    ||||| || ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 781 TGTGATGTCACATTCACACGACCTCGCTGAGCTTTCACAACTTCGACAAATTCG 840
    ||||| || ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 898 CATCTGAGCTTCGAGAGCTGATTTTCTGTGACGCGAGCGGAAATCTTGATTCCT 957
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 841 CATCTGAGCTTCGAGAGCTGATTTTCTGTGACGCGAGCGGAAATCTTGATTCCT 900
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 958 CAAATGACGCGGTTACCTTCTGTAACATTCAAACCTTTTGACTTCTATATGAGCT 1017
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 901 CAAATGACGCGGTTACCTTCTGTAACATTCAAACCTTTTGACTTCTATATGAGCT 960
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1018 GACGTCAGTGTATACCCAAAGCAAGAAAGGGAATAGAAACCTCCAAAGAGTCCAGAT 1077
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 961 GACATCAGTGTATACCCAAAGCAAGAAAGGGAATAGAAACCTCCAAAGAGTCCAGAT 1020
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1078 GATGATCTGCTTCAAGAAATTCAGCAATTTGTATATAAATTTGGAATATGGAAG 1137
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1021 GCTGATATGCTCTTCAGAAATTCGACATCTATCAGAAATTTCTGGAACATTTGAAG 1080
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1138 AATGAAACACTGAAAGCTTACCAATGATGTTCTTCAAAACCAAGATTTCTCGAGAT 1197
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1081 AATGAAACACTGAAAGCTTACCAATGATGTTCTTCAAAACCAAGATTTCTCGAGAT 1140
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1198 TTTTCTTAACCTGGGCTTCTCTGAGAGAGTGAACAAGAACTTGAGAGAAATAGCT 1257
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1141 GTTCGCAAGCTTGGTGTCTCTGAGAGAGTGAACAAGAACTTGAGAGAAATAGCT 1200
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1258 GGCAGTATGAGCTGCTCCTCAAAATCTGATGATGAGGTTGTCTCTCTTTCTTCTGT 1317
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1201 GGCAGTATGAGCTGCTCCTCAAAATCTGATGATGAGGTTGTCTCTCTTTCTTCTGT 1260
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1318 TCTGCGGCTCCAGAACTGATGCTTAAAGGCTCTATTCAGAGAGCGGTAGGTCTCTCTTA 1377
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1261 TCTGCGGCTCCAGAACTGATGCTTAAAGGCTCTATTCAGAGAGCGGTAGGTCTCTCTTA 1320
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1378 CCAGTTTGTGGGCGCCACCCGCTACAGAGCAATGCTGAACGCTGTAATCTGTTTCCCGT 1437
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1321 TCAGCTTGTGGGCTGACCCCAATACAGCGCATGCTCAATGCTGTAATCTGTTTCCCGT 1380
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1438 GCTGAGCTGTCACAGCAAGAGAGAGTCAAGAGAGGCTCTTACCAAGAAATGAGAGATGAT 1497
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1381 GCTGAGAGCCGTCACAGCAAGAGAGAGTCAAGAGAGGCTCTTACCAAGAAATGAGAGATGAT 1440
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1498 -GCTGAGCTGTCACAGCAAGAGAGAGTCAAGAGAGGCTCTTACCAAGAAATGAGAGATGAT 1557
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1441 GCTGAGCTGTCACAGCAAGAGAGAGTCAAGAGAGGCTCTTACCAAGAAATGAGAGATGAT 1500
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1558 TGTGCTGACTGAGAGCTGGGCTTACATCTGAGGTTGCCCCGATTTGGATATTTTCTGCG 1617
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1501 TGTGCTGACTGAGAGCTGGGCTTACATCTGAGGTTGCCCCGATTTGGATATTTTCTGCG 1560
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1618 TGTGCTGACTGAGAGCTGGGCTTACATCTGAGGTTGCCCCGATTTGGATATTTTCTGCG 1677
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1561 TGTGCTGACTGAGAGCTGGGCTTACATCTGAGGTTGCCCCGATTTGGATATTTTCTGCG 1620
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1678 CCAAAATCCAAATATAGTCAACAGTGTGCTGCTGTCATTCGAGGAGGCTGCTGT 1737
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1621 CCAAAATCCAAATATAGTCAACAGTGTGCTGCTGTCATTCGAGGAGGCTGCTGT 1680
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1738 TGTGATGCTGATCTTCAGAAACATACAAAGAAATGCGCATTCCTTTTAAATCATCAGTTA 1797
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1681 TGTGATGCTGATCTTCAGAAACATACAAAGAAATGCGCATTCCTTTTAAATCATCAGTTA 1740
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1798 TTTTCAATTCACAAATATGCTGTTGAGATTTCTGTGATAGTTTAAATGCGGCTGAA 1857
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1741 TTTTCAATTCACAAATATGCTGTTGAGATTTCTGTGATAGTTTAAATGCGGCTGAA 1800
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1858 TTTTCAATTCACAAATATGCTGTTGAGATTTCTGTGATAGTTTAAATGCGGCTGAA 1917
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1801 CTTGACTGTTGTGTGATTCACACACCTCTATCTTAATACACCGATGTCGCGCATCACCCA 1860
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1918 AGGAATTCATTCATGAGAAACCTGCCAGGTGCAACATCTAGATTCACAAATGAACTT 1977
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

```

```

DB 1861 AGGGGTCCAGTTTCATCGAGAGAAACCGCCAGAGTGTCTACATTCAGATTCACGCAAACTT 1920
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 1978 TCTGATTTGATTCATTTATTCACGCTTTGTCTATCCAGAAATAGTTGTTTCAAAAT 2037
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1921 CCTCATCTATATGAGGTTTATTCACGCTTGTCTATCCAGAAATAGTATTTTAAAGT 1980
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
OY 2038 AAGGATCATCTCATTCAGAGTGTGAAAGCAATGCGGAGAAATGSAAGTGT 2091
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
DB 1981 CAGGCACTTCCATTCAGATAGTTTAAGATGACAGCGCAAGAAAGGTTAATG 2034
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

```

```

RESULT 7
AAD48880
ID AAD48880 standard; DNA; 1959 BP.
XX
AC AAD48880;
XX
DT 24-MAR-2003 (first entry)
XX
DE Mouse ABCG5 DNA.
XX
KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW sitosterolemia; hyperlipidaemia; hypercholesterolemia; gall stone;
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW mouse; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
KW ABCG5; gene; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1591
FT /tag="a
FT /product="mABCG5 protein"
XX
PN WO200281691-A2.
XX
PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-US43823.
XX
PR 20-NOV-2000; 2000US-252235P.
XX
PR 28-NOV-2000; 2000US-253645P.
XX
PA (TULU-) TULARK INC.
PA (TEXA) UNTV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
XX
DR WPI: 2003-058548/05.
DR P-PSDB: AAB31702.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating
PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies -
XX
PS Claim 11; Page 73; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
CC invention are useful for treating or preventing sterol-related disorders
CC such as sitosterolemia, hyperlipidaemia, hypercholesterolemia, gall
CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC They are also useful in gene therapy. The present sequence is mouse
CC ABCG5 DNA.
XX
SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 other;

```

```

Query Match 58.4%; Score 1365.4; DB 25; Length 1959;
Best Local Similarity 81.4%; Pred. No. 0;

```


[illegible]

Dp	301	TTCCGGGAGGCTGCGCGGCACTGGGAGCCTGGAGAGGGAGGCTGTTGTGAATGGCTGGCAG	360
Oy	464	CTGCGCGGGAGAGCAAGTTCTCAGAGACTGCTTCTCACTCCTGACAGAGCAGACCCCTGCTG	523
Dp	361	CTGCGGAGGAGCAGCTTTCAGAGACTGCTTCTCACTCCTGACAGAGGAGGAGCTTTTCTCG	420
Oy	524	AGCAGCCTCACCCTGCGGAGACGCTGCATCAACCCGGCTGCTGGCCATCCCGCGGC	583
Dp	421	AGCAGCCTCACTGCGGAGACGCTTGGATATACAGAGGATGCTGGCCCTCTGCGCGAGC	480
Oy	584	AATCCGCGCTCTTCCAGAAAGGTGAGGCGCGATGGGAGACTGATGCTGATGAGCAT	643
Dp	481	TCGCGGCACTTCAACACMAAGATGAGGCAAGTACAGACAGACTGAGCTTGAGCCAC	540
Oy	644	GTCGACAGACCGACTGATTTGGCACTACAGCTTGGGGGCACTTTCACGGGTGAGCGGCG	703
Dp	541	GTCGCGGACCAATATGATTTGGCAGCTATATTTTGGGGCAATTTCCAGTGGCAGCGGCG	600
Oy	704	CGGCTCTCATCCGACGCCAGCTGCTCCAGATCTTAAGTCATGCTGTTTGATGAGCCA	763
Dp	601	CGAATTTCATCCGAGGCCCAACTCTTCAGGAGCCCAAGGTCATGATCTGATGAGCCA	660
Oy	764	ACCAAGGCGCTGGAGCTGCATGACGCTAATGAGATTTGCTGCTCCTGGTGGAACTGGCT	823
Dp	661	ACCAAGGAGCTGGAGCTGCATGACGCAAAATTAATTTGCTTCTTTGGCTGAGAGCTGCT	720
Oy	824	CGCAGGAGCGAATTTGTTGTTCTCAACATTCAACAGCCCGCTGCTGAGCTTTTCAAGTC	883
Dp	721	CGCAGGAGCGAATTTGATTGTATTCACCATCCACAGCCTCGCTTGAAGCTTTTCAACAC	780
Oy	884	TTTGACAAAATTTGCCATCTGAGCTTGGAGAGCTGATTTTCTGTGACGCCACAGCGAA	943
Dp	781	TTTGACAAAATTTGCCATCTGACTTACGAGAGATTTGGTGTCTGTGGGACCCAGAGGAG	840
Oy	944	ATGCTTGATTTCTTCATAGACTGGGGTTTACCCTTGTCTGTAACATTCAACCCTTTTCAC	1003
Dp	841	ATGCTTGCTGCTTTCATTAATCACTTGGTTTACCCTGTCTGTAACATTCAACCCCTTTGAT	900
Oy	1004	TTCTAATTTGAGACCTGAGCTGATGATGATACCCAAAGCAAGAAAGGGAATTTGAATCCCTC	1063
Dp	901	TTTTTACATGGACTTGACATCAGTGTGACACCCAAAGCAGAGCGGGAATTTGAATGCTAC	960
Oy	1064	AAGAGACTCAGATGATAGAACTGCTCTACAGAAATCAGCAATTTGTCAATAAACTTTC	1123
Dp	961	AACGAGTACAGATGCTGGAATGTGCTTCAAGAAATCTGACATCTATCACAAAATCTG	1020
Oy	1124	AAGAAATTTGAAGAAATGAAGAACCTGTAACAACTTACCAATGTTCTTTAAACCAA	1183
Dp	1021	GAGAACTTGAAGAGCAGATTAATCTTAAAACTTACCACAGGTTCTTTTAAAAAATA	1080
Oy	1184	GATTCTCTGAGATTTCTCTAAACTGGGTGTTCTCTGAGAGAGTACACAAAGAACTTG	1243
Dp	1081	GATCTCTCTGGAAATTTCCGGCAACGCTTGTGTCTGCTGAGGCGCAGTACAAAGAACTTA	1140
Oy	1244	GTCAGAAATTAAGCTGCGAGTGAATTCAGCTGCTCTTCAAGATCTGATCATGGTTTGTC	1303
Dp	1141	ATGAGGAATTAAGCAGCAGATGATTAATCTGCTCTCTGAGAAATCTGATCATGGGCTTTC	1200
Oy	1304	CTCCTTTTCTGCTGCGGGGCTCCGAAGCAATGCTTAAAGGTTGCTATCCAGAGCCGC	1363
Dp	1201	CTCATTTTCTACCTTCTCCGCTGCCACAAACACGCTTAAAGGGCGTGTGACAGACCGC	1260
Oy	1364	GTAAGTCTCCTTTTACAGATTTGTGGGGGCAACCCGCTACACAGGCAATGCTAACAGCTGTG	1423
Dp	1261	GTTGGGGGTGCTTATATCAAGTTGTGGGTGCCACCCCATACACCGGCAATGCTTAATGCTGTG	1320
Oy	1424	AATCTGTTTCCGCTGCTGCGAGCTGTACAGCAGCAGAGAGATGAGAGGCGCTTACAG	1483
Dp	1321	AATCTGTTTCCGCTGCTGAGAGCGCTTACAGCAGCAGAGAGATGAGAGGCGCTTATCAT	1380
Oy	1484	AATGAGAGATGATGCGGCTTATGACATGCACTGACGCTCTCCCTTCAAGCTTTGTGGCAC	1543
Dp	1381	AATGAGAGATGCTGCTGCGCTACAGCTTACAGGCTCTCCCTTCAAGCTTATGCGACG	1440

OY	1544	TCGATTTACAGAGTGTGGCTACGACGCTGGGCTTACATCCTGAGGTGCCGATT	1603
Db	1441	GTCAATTTACAGAGTGTGTATTGACTGGCGCTGTATCCGAAAGTGGCCAGATT	1500
OY	1604	GGATATTTTCTGCTGCTCTTGGCCCCCACCATTAAATGGTGAATTTCTACTCTGTG	1663
Db	1501	GGATATTTCTCTGGCTCTTGTGGCCCCCACCATTAAATGGAGATTTCTAACCTGTG	1560
OY	1664	CTACTTGGTATCGTCCAAAATCCAAATATATGACACAGTGTAGTGGCTCTGCTCCATT	1723
Db	1561	CTGCTGGTATAGTCCAAAACCCCTAAATATTTCTCAACAGTATATGGGCTCTGCTAGATC	1620
OY	1724	GGGGGGGTGCTGTGTGGATCTGSAATTCCTCAGAAACATCAAGAAATGCCATTCTTTT	1783
Db	1621	TCGTGGCGCTTATTTGGATCTGGATTTATCGAAGAAACATCAAGAAATGCCATTCTTTA	1680
OY	1784	AAAATCATCACTTATTTTACATTCGCAAAAATATTTGACGTGAGATTCCTGTAGTCATGAG	1843
Db	1681	AAAATCCCTGGGTATTTTATTCATTCGCAAAAATATCTGTGTGAGATTCCTGTGTCAATGAG	1740
OY	1844	TTTACAGGACATTCATTTACATTGTGGAGGCTCAATGTGTTGTGACAACTAATCCATG	1903
Db	1741	TTTTACGGCCCTGAACCTTACCTTGTGTGGATTCACAAACCTTATATGCTAAATCACCCGATG	1800
OY	1904	TGTGCTTTCACCTCAGGAAATTCATTTGATTTGAGAAAACTGGCCGAGTGCAAACTCTAGA	1963
Db	1801	TGCGCCATTCACCCAGAGGGGTCCAGTTCATGAGAAAACTGCCGAGGTGCATATCCAGA	1860
OY	1964	TTTCACATGAACCTTTCGATTTTGTATTTCATTTATTCAGAGCTCTTGTCATCTTAG	2018
Db	1861	TTTCAGCGCAACCTTCATCTTATTTATTTATTTATTTTCCAGAGCTCTTGTCATCTTAG	1915

RESULT 10	ID	ABK51687	standard; cDNA; 1069 BP.
XX	AC	ABK51687;	
XX	DT	30-JUL-2002	(first entry)
XX	DE	cDNA encoding hamster ABCG5 protein.	
XX	HM	hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol,	
KM	arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;		
ss.			
XX	OS	Cricetinae sp.	
XX	Key	Location/Qualifiers	
FH	CDS	30..1049	
FT		/tag= a	
FT		/partial	
FT		/product= "Hamster ABCG5 protein"	
FT		/note= "This sequence lacks both a start and stop codon"	
PN	WO200227016-A2.		
XX			
PD	04-APR-2002.		
XX			
PF	25-SEP-2001; 2001WO-US29859.		
XX			
PR	25-SEP-2000; 2000US-235268P.		
XX			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
PA	(PATE/) PATEL S B.		
PA	(DEAN/) DEAN M.		
XX			
PI	Patel SB, Dean M;		
XX			
WPI	2002-416483/44.		
DR	P-PsDB; AAU96987.		

xx Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 pt nucleic acid encoding the polypeptide, useful for treating
 pt sitosterolemia, arteriosclerosis and heart diseases

xx Example 3; Page 47; 66pp; English.

cc The present invention relates to a new mammalian ATP-binding cassette
 cc gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 cc predisposition for developing sitosterolemia, arteriosclerosis or heart
 cc disease. The molecules of the invention are also useful for identifying
 cc a compound which alters ABCG5 activity level comprising contacting a cell
 cc culture or mammal which have ABCG5 polypeptide with a compound and
 cc measuring ABCG5 biological activity in the cell culture or in mammal,
 cc where an increase or decrease in ABCG5 biological activity compared to
 cc ABCG5 biological activity in a control cell culture or mammal not
 cc contacted with the compound, identifies a compound that increases or
 cc decreases ABCG5 activity respectively. The cell culture or mammal
 cc comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 cc ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 cc polypeptide in a cell culture or mammal is also compared with that of a
 cc second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 cc Stimulation of ABCG5 activity is useful for treating or preventing
 cc hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 cc disease. The method of the invention is useful for increasing cholesterol
 cc excretion and/or decreasing cholesterol adsorption. The present nucleic
 cc acid sequence encodes the hamster ABCG5 protein of the invention.

xx Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 other;

Query Match 32.9%; Score 770.2; DB 24; Length 1069;
 Best Local Similarity 83.7%; Pred. No. 2.2e-203;
 Matches 896; Conservative 0; Mismatches 173; Indels 2; Gaps 2;

QY 368 TCAGGCTCCGGGAAACACGCTGCTGAGCCCATGTCGGGAGCGGGCGCGCGG 427
 Db 1 TCAGGCTCCGGGAAACACGCTGCTGAGCCCATGTCGGGAGCGGGCGCGCGG 59
 QY 428 ACCTTCTGGGGAGGTGTATGTGAACGCGCGCGCGCGCGCGGAGATTCCAGGAC 487
 Db 60 ACCCTGGAGGGGGAGGTGTGTGAACGCGCGCGCGCGCGGAGATTCCAGGAC 119
 QY 488 TGGTTCCTCTACCTCTGCGAGAGGACACCTCTGAGAGCCTCAGCCGCGGAGAGC 547
 Db 120 TGGTTCCTCTATCTCTGCGAGAGGACGCTCTTCTGAGAGCTCAAGGCGGAGAGC 179
 QY 548 CTCACATACACCGCGCTGCGCATCCCGCGCGGGAATCCCGGCTCTCCAGGAAGAG 607
 Db 180 CTCGCTACACGCGCGATGCTGCGCCCTCCGAGTAGTCTTCCGAGCTTCTATGACAGAG 239
 QY 608 GTGAGAGCCCTCATGCGAGAGCTGAGCTGAGCCATGTGCGACAGCAGTATGGCAAC 667
 Db 240 GTAGAGGAGAGTCTATGAGAGCTAAGCTGAGCCAGTGGCAGACAGGATATGGCAAC 299
 QY 668 TACAGCTGGGGGCAATTTCCAGCGGTGAGCGGCGCGGCTCTCATTCGAGCCAGCTG 727
 Db 300 TATTAATTTGGGGAAATTTCCAGTGGCGAGCGCGCGGAGTCCATTCGAGCCAGCTG 359
 QY 728 CTCGAGATCTTAAGGATGCTGTGATGAGCCAAACACAGAGCCTGAGCTCAGTACT 787
 Db 360 ATTTCAGAGACCCCAAGTATGATGATGATGAGCCAAACACAGAGTCTGAGCTCAGTACT 419
 QY 788 GCTAATCATATGCTCTCTCTCTGAGTGAAGTGGCTGCGAGAAACGAATGTGTTCTC 847
 Db 420 GCAATATCAATGTTCTCTCTCTGAGTGAAGTGGCTGCGAGAAACGAATGTGTTCTC 479
 QY 848 ACCATTCACAGAGCCCGCTCTCTCTGAGTGAAGTGGCTGCGAGAAATTCATTCGAGC 907
 Db 480 ACCATTCACAGAGCCCGCTCTCTCTGAGTGAAGTGGCTGCGAGAAATTCATTCGAGC 539
 QY 908 TTGCGAGAGCTGATTTCTGCGAGAGCGAGGAAATCTGATTTCTTCAATGACTGC 967
 Db 540 TACGGAGAGATGCTGTTCTGCGAGAGCGGAGGAAATCTGATTTCTTCAATGACTGC 599

QY 968 GATTACCTCTCTCTGAGTGAAGTGGCTGCGAGAAACGAATGTGTTCTC 1027
 Db 600 GATTACCTCTCTCTGAGTGAAGTGGCTGCGAGAAACGAATGTGTTCTC 659
 QY 1028 GATACCCCAAGAGAGGAAATGAAACCTTCAAGAGAGTCCAGATGATGATCT 1087
 Db 660 GATACCCCAAGAGAGGAAATGAAACCTTCAAGAGAGTCCAGATGATGATCT 719
 QY 1088 GCTTACAGAAATGAGCAATTTCTATTAACCTTGAAGAAATGTAAGAAACAC 1147
 Db 720 GCTTACAGAGACTGTGAGTCTGACAAATCTCGGAGAAATGTAAGAAACAC 779
 QY 1148 CTGAAGAGTGAAGTGTCTCTTCAAAACCAAGATCTCTGAGTGTCTTCA 1207
 Db 780 CTGAAGAGTGAAGTGTCTCTTCAAAACCAAGATCTCTGAGTGTCTTCA 839
 QY 1208 CTGGGTGTCTCTGAGAGAGAGTGAAGAAACCTTGTGAGAAATGAGTGTAT 1267
 Db 840 CTGGGTGTCTCTGAGAGAGAGTGAAGAAACCTTGTGAGAAATGAGTGTAT 899
 QY 1268 ACGGTCTCTCTCAAGATGATGATGATGATGATGATGATGATGATGATG 1327
 Db 900 ATGCGTCTTGTCAAGATGATGATGATGATGATGATGATGATGATGATG 959
 QY 1328 CGAAGCAATGCTTAAGAGGCTATGACAGAGCGCTGAGTGTCTTCA 1387
 Db 960 CAGAGCAATGCTTAAGAGGCTATGACAGAGCGCTGAGTGTCTTCA 1018
 QY 1388 GGCCTCACCCCTGATACAGAGAGTGAAGCTGATGATGATGATGATG 1438
 Db 1019 GCGCCCAACCCGATACAGAGAGTGAAGCTGATGATGATGATGATG 1069

RESULT 11
 AAD22022
 ID AAD22022 standard; DNA; 472 BP.
 XX
 AC AAD22022;
 DT 12-FEB-2002 (first entry)
 XX
 XX Human sitosterolemia susceptibility gene (SSG) exon 13.
 DE
 XX Human; sitosterolemia susceptibility gene; SSG; arteriosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; ds.
 OS
 XX Homo sapiens.
 PN W0200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001MO-US12758.
 XX
 PR 18-APR-2000; 2000US-198465P.
 XX
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 DR WPI; 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 PS
 XX Claim 73; Fig 14B; 105pp; English.

CC The invention relates to an isolated sitosterolaemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is an exon of human SSG DNA.

XX Sequence 472 BP; 134 A; 93 C; 100 G; 145 T; 0 other;

Query Match 20.28; Score 472; DB 24; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.2e-120;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1869 GCAGCTCAATGTTTGTGACAACTAATCCATGTCCTTCACTCAAGAAATCAAT 1928
DB 1 GCAGCTCAATGTTTGTGACAACTAATCCATGTCCTTCACTCAAGAAATCAAT 60
QY 1929 TCATTGAGAAAACCTGCCAGGTGACAACTATGATTCACAAATGACTTTCATTTGT 1988
DB 61 TCATTGAGAAAACCTGCCAGGTGACAACTATGATTCACAAATGACTTTCATTTGT 120
QY 1989 ATTGATTTTTCAGCTCTTGTTCATCTTGAAGAAATGATTTGTTTCAAAATAGGATC 2048
DB 121 ATTGATTTTTCAGCTCTTGTTCATCTTGAAGAAATGATTTGTTTCAAAATAGGATC 180
QY 2049 TCATTGACAGTAGTAGGAAAGCCATGCTGGGAAAATGAAAGCAACCTGCCACTGTGCA 2108
DB 181 TCATTGACAGTAGTAGGAAAGCCATGCTGGGAAAATGAAAGCAACCTGCCACTGTGCA 240
QY 2109 TGACGTCTCTGAACGTCTGAATGAGAGTGCATGATTTCTTTCTTGACAGACATCTC 2168
DB 241 TGACGTCTCTGAACGTCTGAATGAGAGTGCATGATTTCTTTCTTGACAGACATCTC 300
QY 2169 AAGCTTTTAAACATTAAACATTCATTTGCTCTTGATTCACAGAGCCTTGATGC 2228
DB 301 AAGCTTTTAAACATTAAACATTCATTTGCTCTTGATTCACAGAGCCTTGATGC 360
QY 2229 AATGAAAGTGTATTATAGTCCCTGCTTACAACTTGAGAGGACATGCTTATTTGCA 2288
DB 361 AATGAAAGTGTATTATAGTCCCTGCTTACAACTTGAGAGGACATGCTTATTTGCA 420
QY 2289 AATTGTGACTGAGCGGACCAAGAAATGTAATTAATTCATTAACCTATGGG 2340
DB 421 AATTGTGACTGAGCGGACCAAGAAATGTAATTAATTCATTAACCTATGGG 472

RESULT 12
AAC76065
ID AAC76065 standard; cDNA; 432 BP.

XX AAC76065;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1620 polynucleotide sequence SEQ ID NO:1239.

XX Human: open reading frame; ORFX: detection; cytosolic; hepatotropic;
KM vulnery; antiparkinsonian; nootropic; neuroprotective;
KM anticonvulsant; osteoparalytic; antidiarrhoeal; immunosuppressant; cardiant;
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
KM antiviral; antibacterial; antifungal; antipneumatic; antithyroid;
KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM cholesterol ester storage; systemic lupus erythematosus; infection;
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM bone damage; cartilage damage; antinflammatory disease; coagulation;
KM thrombosis; contraceptive; ss.

XX Homo sapiens.

XX W0200058473-A2.

XX 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkels RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PDB; AAB41856.

XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 5; Page 2444; 5507bp; English.

CC AAC7446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
CC antiparkinsonian; antidiarrhoeal; nootropic; neuroprotective;
CC osteoparalytic; anticonvulsant; antidiarrhoeal; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antinflammatory; antibacterial; antiviral; antifungal; antipneumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 432 BP; 87 A; 110 C; 118 G; 117 T; 0 other;

Query Match 18.48; Score 429.4; DB 21; Length 432;
Best Local Similarity 99.8%; Pred. No. 7.9e-109;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1175 AAACCAAGATTCCTCCGAGTTCCTTAACCTGAGGTCTTCTTCAAGAGAGTACCA 1234
DB 1 AAACCAAGATTCCTCCGAGTTCCTTAACCTGAGGTCTTCTTCAAGAGAGTACCA 60
QY 1235 AGAACTTGAGGAAATAGCTGCACTGATATGCGCTTCCCTTCAAGATTCGATG 1294
DB 61 AGAACTTGAGGAAATAGCTGCACTGATATGCGCTTCCCTTCAAGATTCGATG 120
QY 1295 GGTTCCTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1354
DB 121 GGTTCCTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180

QY 1355 CAGGACCGCGTAGGTCCTTACGAGTTGTGGGCGCACCCGTTACACAGCATGCTG 1414
DB 181 CAGGACCGCGTAGGTCCTTACGAGTTGTGGGCGCACCCGTTACACAGCATGCTG 240
QY 1415 AACGCTGTGAATGTGTTCCCGTGCAGACTGTGTAGCAGCAGAGAGTCAGACGGC 1474
DB 241 AACGCTGTGAATGTGTTCCCGTGCAGACTGTGTAGCAGCAGAGAGTCAGACGGC 300
QY 1475 CTTACACAGAGTGCAGATGATGCTGACCTATGACAGTCCTCTCCCTTACAGCCTT 1534
DB 301 CTTACACAGAGTGCAGATGATGCTGACCTATGACAGTCCTCTCCCTTACAGCCTT 360
QY 1535 GTTGCACCATGATTTTTCAGCAGTGTGTCTACTGACGCTGAGCTTACATCCTGAGTT 1594
DB 361 GTTGCACCATGATTTTTCAGCAGTGTGTCTACTGACGCTGAGCTTACATCCTGAGTT 420
QY 1595 GCCCGATTGGG 1605
DB 421 GCCCGATTGGG 431

RESULT 13

AA294755
ID AA294755 standard; cDNA; 281 BP.

XX AA294755;

DT 01-AUG-2000 (first entry)

DE Human ATP binding cassette cDNA fragment 168043.

KW ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; ss.

OS Homo sapiens.

PN WO200018912-A2.

PD 06-APR-2000.

PF 21-SEP-1999; 99WO-EP06991.

PR 25-SEP-1998; 98US-0101706.

PA (FARB) BAYER AG.

PI Schmitz G, Klucken J;

DR WPI; 2000-293151/25.

PT Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
PS Claim 9; Page 135; 154pp; English.

CC The present sequence is that of human ATP binding cassette
CC (ABC) cDNA fragment 168043, identified as a cholesterol-sensitive
CC gene fragment. The invention provides cholesterol-sensitive ABC
CC genes (see AA294734-63). These genes, and polypeptides encoded by
CC them, can be used for diagnostic and therapeutic applications, and
CC for biochemical or cell-based assays to screen for pharmacologically
CC active modulator compounds useful for the treatment of lipid
CC disorders, atherosclerosis or other inflammatory diseases such as
CC psoriasis and lupus erythematosus.

XX Sequence 281 BP; 60 A; 68 C; 73 G; 80 T; 0 other;

Query Match 11.5%; Score 268; DB 21; Length 281;
Best Local Similarity 99.6%; Pred. No. 4.8e-64;
Matches 279; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1175 AAACCAAGATTTCTCGTAGGTTTCTCTAAACGTGGTGTCTCTGAGAGAG--TGAC 1233
DB 1 AAACCAAGATTTCTCGTAGGTTTCTCTAAACGTGGTGTCTCTGAGAGAGTTGAC 60
QY 1234 AAGAACTGTTGGTGAATAATAGCTGACAGTGTATACGCTCTCCCTTACAAATCTGATCAT 1293
DB 61 AAGAACTGTTGGTGAATAATAGCTGACAGTGTATACGCTCTCCCTTACAAATCTGATCAT 120
QY 1294 GGGTTGTCTCTCTCTTTTCTTGTGTTGCGGGTCCGAGCAATGTCTAAAGGTGCTAT 1353
DB 121 GGGTTGTCTCTCTCTTTTCTTGTGTTGCGGGTCCGAGCAATGTCTAAAGGTGCTAT 180
QY 1354 CCAGGACCGCGTAGGTCCTCTTACAGATTTGTGGGCGCACCCGTTACAGAGCATGCT 1413
DB 181 CCAGGACCGCGTAGGTCCTCTTACAGATTTGTGGGCGCACCCGTTACAGAGCATGCT 240
QY 1414 GAACGCTGTGAATCTGTTTCCCGTGTGCGACGCTGTACG 1453
DB 241 GAACGCTGTGAATCTGTTTCCCGTGTGCGACGCTGTACG 280

RESULT 14

ABK51683
ID ABK51683 standard; DNA; 5460 BP.

XX ABK51683;

DT 30-JUL-2002 (first entry)

DE Human ABCG5 upstream genomic sequence, exon 1, intron 1 and exon 2.

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW chromosome 2p21; ds.

OS Homo sapiens.

PN WO200227016-A2.

PD 04-APR-2002.

PF 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

PI Patel SB, Dean M;

DR WPI; 2002-416483/44.

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases -
PS Example 3; Page 38-41; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a

CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the upstream genomic sequence, exon 1, intron 1
CC and exon 2 of the human ABCG5 gene located on chromosome 2p21.
XX

Sequence 5460 BP; 1351 A; 1350 C; 1508 G; 1243 T; 8 other;

Query Match 10.7%; Score 249.6; DB 24; Length 5460;
Best Local Similarity 98.4%; Pred. No. 3.8e-58;
Matches 252; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGAGGCTCCCACTGAGCCACTCTGGGGA 60
DB 4504 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGAGGCTCCCACTGAGCCACTCTGGGGA 4563
OY 61 GGGTCGGCCACACAGAAATTTGGCCAGCTTGGCTGTGGCCATGGGTGACCTCTC 120
DB 4564 GGGTCGGCCACACAGAAATTTGGCCAGCTTGGCTGTGGCCATGGGTGACCTCTC 4623
OY 121 ATCTTTGACCCCGGAGGAGTCCATGGGTCTCCAGTAACAGAGGCTCCAGAGCTCCCT 180
DB 4624 ATCTTTGACCCCGGAGGAGTCCATGGGTCTCCAGTAACAGAGGCTCCAGAGCTCCCT 4683
OY 181 GGAGGGGCTCTCTGCTCCAGCCCGGAGGCTCCAGAGCTCCAGAGCTCCCTA 240
DB 4684 GGAGGGGCTCTCTGCTCCAGCCCGGAGGCTCCAGAGCTCCAGAGCTCCCTA 4743
OY 241 CAGCGTCAGCCAGCCG 256
DB 4744 CAGCGTCAGGTAAGCC 4755

RESULT 15

AAD22010
ID AAD22010 standard; DNA; 249 BP.

XX AAD22010;

XX 12-FEB-2002 (first entry)

XX Human sitosterolemia susceptibility gene (SSG) exon 1.

KW Human; sitosterolemia susceptibility gene; SSG; arteriosclerosis;
KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
KW xanthoma; haemolytic anaemia; transgenic animal; ds.

XX Homo sapiens.

XX WO200179272-A2.

XX 25-OCT-2001.

XX 18-APR-2001; 2001WO-US12758.

XX 18-APR-2000; 2000US-198465P.

XX 15-MAY-2000; 2000US-204234P.

XX (TULA-) TULARIK INC.

XX Tian H, Schultz J, Shan B;

XX WPI; 2002-017598/02.

XX Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder

PS Claim 73; Fig 14B; 105pp; English.

XX The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, arteriosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is an exon of human SSG DNA.

Sequence 249 BP; 44 A; 86 C; 74 G; 45 T; 0 other;

Query Match 10.6%; Score 249; DB 24; Length 249;
Best Local Similarity 100.0%; Pred. No. 8.6e-59;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGAGGCTCCCACTGAGCCACTCTGGGGA 60
DB 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGAGGCTCCCACTGAGCCACTCTGGGGA 60
OY 61 GGGTCGGCCACACAGAAATTTGGCCAGCTTGGCTGTGGCCATGGGTGACCTCTC 120
DB 61 GGGTCGGCCACACAGAAATTTGGCCAGCTTGGCTGTGGCCATGGGTGACCTCTC 120
OY 121 ATCTTTGACCCCGGAGGAGTCCATGGGTCTCCAGTAACAGAGGCTCCAGAGCTCCCT 180
DB 121 ATCTTTGACCCCGGAGGAGTCCATGGGTCTCCAGTAACAGAGGCTCCAGAGCTCCCT 180
OY 181 GGAGGGGCTCTCTGCTCCAGCCCGGAGGCTCCAGAGCTCCAGAGCTCCCTA 240
DB 181 GGAGGGGCTCTCTGCTCCAGCCCGGAGGCTCCAGAGCTCCAGAGCTCCCTA 240
OY 241 CAGCGTCAG 249
DB 241 CAGCGTCAG 249

Search completed: July 27, 2003, 00:55:32
Job time : 570.197 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)

11512.153 Million cell updates/sec

.....atatcataaacctatg 2340

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AX320883	AX320883	Sequence 4 from Patent WO0179272.	AX320883	AX320883.1	GI:17902433	Homo sapiens (human)	Homo sapiens	1	Tian, H., Schultz, J. and Shan, B.	Sitosterolemia susceptibility gene (ssg): compositions and methods of use

JOURNAL Patent: WO 0179272-A 4 25-OCT-2001;
Tularik Inc. (US)
FEATURES Location/Qualifiers
Source 1. 2340
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="human sitosterolemia gene (SSG)"
107..2062
protein
/note="human sitosterolemia susceptibility gene (SSG)
protein"
/protein_id="CAD19409.1"
/db_xref="GI:17902434"
/translation="MGDLSLTPGSGMGLQVNRGSQSLEGAAPAPAPEPHSLGILHAS
YSVSHVRPMDITSCROOITROILKDVSLYVESQIIMCIGSSGSKTLLDAMSGR
LGRAGTFLFEVYNGRRLRREPODFSVLOSDDLSSLYETHTAYALLARGN
PGSFQKKEVAVMAELSLSHVADRLIGNVSLGGISGERRVSIAAOLDDPKYLMDE
PTTGDCMTANOIVYLVELARNRKLYVTITHPRSELQLEPKIALISFELFEGC
PAMLPDFNDGCPPEHNPEDFYMDLTSVDQSKEREIETSKRVOMTESAVKKSAI
CHRTLNIEBMRHLKTLPMVPEKTDSPGVSKLGLYLRVRNLRNLAIVTLILQ
NLIMGLFELFVLRVSNVLKGIODRVLIOFVGATPYTGMLANVLPVLRVSD
QESODGLYKRMOMLALYALHVLPEFVYVATMIFSVCTYMLGLHPVARFGPSAALLA
PHLIGELTLVLIGLIVQNPINVSVALISIAVGSLPRIDQMPYPRKISITFT
FOKYSSEILVINEFVGLNFTCCSSNYSVTTPNCAFTQGIQIFLEKTPCATSRFTNMF
LIIKSTPILVILIGLIVETKIRDLISR"

BASE COUNT 541 a 601 c 598 g 600 t
ORIGIN

Query Match 100.0%; Score 2340; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCAGGTGAGCAGCAGGAGCAGTCCACAGGGGCTCCCAACTGAAGCAGCTGGGG 60
DB 1 GTCAGGTGAGCAGCAGGAGCAGTCCACAGGGGCTCCCAACTGAAGCAGCTGGGG 60
QY 61 GGGTCCGGCCACAGAAAATTGGCCAGCTTGTGCTGCTGGCCATGGGTGACCTCTC 120
DB 61 GGGTCCGGCCACAGAAAATTGGCCAGCTTGTGCTGCTGGCCATGGGTGACCTCTC 120
QY 121 ATCTTGACCCCGGAGGGGTCATGGTCCAGAAAGACAGAGTCCCAAGAGCTCCCT 180
DB 121 ATCTTGACCCCGGAGGGTCCATGGGTCCAGAAAGACAGAGTCCCAAGAGCTCCCT 180
QY 122 ATCTTGACCCCGGAGGGTCCATGGGTCCAGAAAGACAGAGTCCCAAGAGCTCCCT 180
DB 122 ATCTTGACCCCGGAGGGTCCATGGGTCCAGAAAGACAGAGTCCCAAGAGCTCCCT 180
QY 181 GAGAGGGGCTCTGCCACCGCCCGGAGGCTCACAAGCTGGGCAATCCTCATGCTCTTA 240
DB 181 GAGAGGGGCTCTGCCACCGCCCGGAGGCTCACAAGCTGGGCAATCCTCATGCTCTTA 240
QY 241 CAGCTCAGCAGCCCGGTAGAGGCTGTGGGACATCATCTTGGCCGAGCAGATGGAC 300
DB 241 CAGCTCAGCAGCCCGGTAGAGGCTGTGGGACATCATCTTGGCCGAGCAGATGGAC 300
QY 242 CAGCTCAGCAGCCCGGTAGAGGCTGTGGGACATCATCTTGGCCGAGCAGATGGAC 300
DB 242 CAGCTCAGCAGCCCGGTAGAGGCTGTGGGACATCATCTTGGCCGAGCAGATGGAC 300
QY 301 CAGCAGATCTCAAGATGTCTCTTACGTGAGGAGAGCGGGAGATCATGTGATCTCT 360
DB 301 CAGCAGATCTCAAGATGTCTCTTACGTGAGGAGAGCGGGAGATCATGTGATCTCT 360
QY 302 CAGCAGATCTCAAGATGTCTCTTACGTGAGGAGAGCGGGAGATCATGTGATCTCT 360
DB 302 CAGCAGATCTCAAGATGTCTCTTACGTGAGGAGAGCGGGAGATCATGTGATCTCT 360
QY 361 AGGAAGCTCAGGCTCCGGGAAAACACAGCTGTGAGAGCCATGTCCGGAGGCTGGGG 420
DB 361 AGGAAGCTCAGGCTCCGGGAAAACACAGCTGTGAGAGCCATGTCCGGAGGCTGGGG 420
QY 421 CGGGGGAGCTTCTCTGGGGAGGTGTATGTAAAGCGCGGCGCTGCCGGAGACAGTT 480
DB 421 CGGGGGAGCTTCTCTGGGGAGGTGTATGTAAAGCGCGGCGCTGCCGGAGACAGTT 480
QY 481 CCAGAGCTCTTCTCTACGTCTGAGAGGAGACACCTGTGAGACAGCTCAACCTGGG 540
DB 481 CCAGAGCTCTTCTCTACGTCTGAGAGGAGACACCTGTGAGACAGCTCAACCTGGG 540
QY 541 CGAGAGCTGACACTACACCGCGCTGTGCGCAATCCGCGGCAATCCGCGCTCTTCCA 600
DB 541 CGAGAGCTGACACTACACCGCGCTGTGCGCAATCCGCGGCAATCCGCGCTCTTCCA 600

QY 601 GAAGAAGGTGAGCCGCTCATGTGACAGAGCTGAGTGCATGTGGCAGACCGACTGAT 660
DB 601 GAAGAAGGTGAGCCGCTCATGTGACAGAGCTGAGTGCATGTGGCAGACCGACTGAT 660
QY 661 TGGCAACTACAGCTTGGGGGGCATTTTCACAGGGTGAGCGCGCGGCTCCATCCGAC 720
DB 661 TGGCAACTACAGCTTGGGGGGCATTTTCACAGGGTGAGCGCGCGGCTCCATCCGAC 720
QY 721 CCAGCGCTCCAGAGCTCAAGTCAATGCTGTTGATGAGCCACACAGAGCTGACAC 780
DB 721 CCAGCGCTCCAGAGCTCAAGTCAATGCTGTTGATGAGCCACACAGAGCTGACAC 780
QY 781 CATGACTGTATATCAGATTTGCTCTCTGCTGAGTGAACAGGCTCGCAGAACCGAATTTG 840
DB 781 CATGACTGTATATCAGATTTGCTCTCTGCTGAGTGAACAGGCTCGCAGAACCGAATTTG 840
QY 841 GGTTCACACATTCACACAGCCCGCTGTGAGCTTTTTCAGCTTTTTCAGAAAATTCAT 900
DB 841 GGTTCACACATTCACACAGCCCGCTGTGAGCTTTTTCAGCTTTTTCAGAAAATTCAT 900
QY 901 CCTGACCTTCGGAGAGCTATTTCTGTGGCACCGCAGGGAATGCTGATTTCTTCA 960
DB 901 CCTGACCTTCGGAGAGCTATTTCTGTGGCACCGCAGGGAATGCTGATTTCTTCA 960
QY 961 TGACTCGGTTACCTTGTCTGACATTCAAAACCTTTTGAATCTTATATGAGACTGAC 1020
DB 961 TGACTCGGTTACCTTGTCTGACATTCAAAACCTTTTGAATCTTATATGAGACTGAC 1020
QY 1021 GTCAGTGTATACCAAGACAGAGACGGAATAGAAACCTCCAGAGAGTCCAGATGAT 1080
DB 1021 GTCAGTGTATACCAAGACAGAGACGGAATAGAAACCTCCAGAGAGTCCAGATGAT 1080
QY 1081 AGAATCTGCTCAACAAATTCAGCAATTTGCTATAAACTTTGAATAATTTGAAAGAT 1140
DB 1081 AGAATCTGCTCAACAAATTCAGCAATTTGCTATAAACTTTGAATAATTTGAAAGAT 1140
QY 1141 GAAACACCTGAAAACCTTACCAATGCTTCTTCAAAACCAAGATTCCTGAGATTT 1200
DB 1141 GAAACACCTGAAAACCTTACCAATGCTTCTTCAAAACCAAGATTCCTGAGATTT 1200
QY 1201 CTCTAATCTGGGTCTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGCTGGC 1260
DB 1201 CTCTAATCTGGGTCTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGCTGGC 1260
QY 1261 AGTGAATACGCTCCCTCAAAATCTGAATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 AGTGAATACGCTCCCTCAAAATCTGAATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 GCGGGTCCGAAACAAATGTGCTAAAGGCTGTATCCAGAGACGCGTATGCTCTTACCA 1380
DB 1321 GCGGGTCCGAAACAAATGTGCTAAAGGCTGTATCCAGAGACGCGTATGCTCTTACCA 1380
QY 1381 GTTGTGGGCGCCACCCGCTACACAGGCAAGCTGTGAATCTGTGATCTGCTGCT 1440
DB 1381 GTTGTGGGCGCCACCCGCTACACAGGCAAGCTGTGAATCTGTGATCTGCTGCT 1440
QY 1441 GCGAGCTGCAGCAGACAGAGAGTCAAGAGCGCTTACCAAGAGTGGCAGATATATGCT 1500
DB 1441 GCGAGCTGCAGCAGACAGAGAGTCAAGAGCGCTTACCAAGAGTGGCAGATATATGCT 1500
QY 1501 GGCCTATGACACTGACAGCTCTCCCTTACAGCTTGTTCACACATGATTTTCAGAGTGT 1560
DB 1501 GGCCTATGACACTGACAGCTCTCCCTTACAGCTTGTTCACACATGATTTTCAGAGTGT 1560
QY 1561 GTGCTACTGAGAGCGTGGGCTTACATCTGAGAGTTCGCGAATTTGATTTTTCGTGTC 1620
DB 1561 GTGCTACTGAGAGCGTGGGCTTACATCTGAGAGTTCGCGAATTTTTCGTGTC 1620
QY 1621 TCTCTTGGCCCCCACTTAATTTGAGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 TCTCTTGGCCCCCACTTAATTTGAGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1681 AAATCAAAATATGACAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

OY	841	GGTTTCTCACACATTTACACACAGCCCGGTTCTGAGCTTTTTCAGCTCTTTTGACAAAAATTCACAT	900
Db	841	GGTTTCTCACACATTTACACACAGCCCGGTTCTGAGCTTTTTCAGCTCTTTTGACAAAAATTCACAT	900
OY	901	CCTGAGCTTCGGAGAGAGCTGATTTCTGTGGGACAGCGAGGGAAATGCTTGATTTCTTCAA	960
Db	901	CCTGAGCTTCGGAGAGAGCTGATTTCTGTGGGACAGCGAGGGAAATGCTTGATTTCTTCAA	960
OY	961	TGACTGCGGTTACCCCTTGCTCTGAAACATTCAAACCCCTTTTGACTTCTATATGAGCTGCAC	1020
Db	961	TGACTGCGGTTACCCCTTGCTCTGAAACATTCAAACCCCTTTTGACTTCTATATGAGCTGCAC	1020
OY	1021	GTCACTGGATACCCCAAGCAGGAAGGAAATAGAAACCTTCAGAGAGCTCAGATGAT	1080
Db	1021	GTCACTGGATACCCCAAGCAGGAAGGAAATAGAAACCTTCAGAGAGCTCAGATGAT	1080
OY	1081	AGAACTCGCTACAGAAATACGCAATTTGTCATATAACTTTGAAAGATATTGAAAGAT	1140
Db	1081	AGAACTCGCTACAGAAATACGCAATTTGTCATATAACTTTGAAAGATATTGAAAGAT	1140
OY	1141	GAAACACCTGAAAGAGTTACCAATGGTTCCTTCAAAACCAAGAAATTCCTCCGGAGTTT	1200
Db	1141	GAAACACCTGAAAGAGTTACCAATGGTTCCTTCAAAACCAAGAAATTCCTCCGGAGTTT	1200
OY	1201	CTCTAACTGGGTTCTCTCTGAGAGAGGTGACAAAGAACTGGTGAAGAAATTAACCTGGC	1260
Db	1201	CTCTAACTGGGTTCTCTCTGAGAGAGGTGACAAAGAACTGGTGAAGAAATTAACCTGGC	1260
OY	1261	AGTGAATTACGGGCTCTCTCTTCAGAAATCTGATCATGCGGTTTGTTCCTCTTCTCTGTTCT	1320
Db	1261	AGTGAATTACGGGCTCTCTCTTCAGAAATCTGATCATGCGGTTTGTTCCTCTTCTCTGTTCT	1320
OY	1321	GGGGTCCGGAAGCAATGTGCTAAAGGGGGCTATCCAGAGACCGCGTAAAGTCTCTTACCA	1380
Db	1321	GGGGTCCGGAAGCAATGTGCTAAAGGGGGCTATCCAGAGACCGCGTAAAGTCTCTTACCA	1380
OY	1381	GTTTGTGGGGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTCT	1440
Db	1381	GTTTGTGGGGCCACCCCGTACACAGGCATGCTGAACGCTGTGAATCTGTTTCCCGTCT	1440
OY	1441	GGGAGCTGTACGCAACCGAGAGAGTCAAGAGAGGCTCTACAGAGAGTGGCAGATGATGCT	1500
Db	1441	GGGAGCTGTACGCAACCGAGAGAGTCAAGAGAGGCTCTACAGAGAGTGGCAGATGATGCT	1500
OY	1501	GGCCCTATGACATGACAGCTCTCCCGCTTACAGGCTGTGGTGGCCACCAATGATTTACAGAGCT	1560
Db	1501	GGCCCTATGACATGACAGCTCTCCCGCTTACAGGCTGTGGTGGCCACCAATGATTTACAGAGCT	1560
OY	1561	GTCGTACTGAGAGCTGGGGCTTACATCTCGAGAGTTGCCAGATTTGGATATTTTCTGCTGC	1620
Db	1561	GTCGTACTGAGAGCTGGGGCTTACATCTCGAGAGTTGCCAGATTTGGATATTTTCTGCTGC	1620
OY	1621	TCTCTTGGCCCCGACATTAATTTGGTGAATTTCTAACTCTTGTGCTACTTGTGATTCGTCCA	1680
Db	1621	TCTCTTGGCCCCGACATTAATTTGGTGAATTTCTAACTCTTGTGCTACTTGTGATTCGTCCA	1680
OY	1681	AAATCCAAATTAATGTCAAACAGTGTAGTGGGCTGCTGCTGCAATTCGGGGGGCTGTGGG	1740
Db	1681	AAATCCAAATTAATGTCAAACAGTGTAGTGGGCTGCTGCTGCAATTCGGGGGGCTGTGGG	1740
OY	1741	ATCTGGAATCTCTCAGAAACATACAGAAATGCCATTCCTTTAAATCATCAATATT	1800
Db	1741	ATCTGGAATCTCTCAGAAACATACAGAAATGCCATTCCTTTAAATCATCAATATT	1800
OY	1801	TACATTTCCAAAATTAATGCAAGTGAATTTCTTTATGTCATAGAGATTTCTACGAGCTGAATTT	1860
Db	1801	TACATTTCCAAAATTAATGCAAGTGAATTTCTTTATGTCATAGAGATTTCTACGAGCTGAATTT	1860
OY	1861	CACCTTGTGAGAGCTCAAAATGTTTCTGTGACAACATTAACCAATGTGCTCTTCACTCAAG	1920
Db	1861	CACCTTGTGAGAGCTCAAAATGTTTCTGTGACAACATTAACCAATGTGCTCTTCACTCAAG	1920
OY	1921	AATTCAATTATGAGAAAACTGCCAGGTGCACAATCTAGATTCAATGAATCTTCT	1980

Db	1921	AATTCATTCAATTGAGAAACCCTGCCAGGTGCACATCTAGATTCACAAAGAACTTCT	1980
QY	1981	GATTTTGTATTCATTATTTCCAGGCTCTGTATCATCTAGAAATAGTGTGTTTCAAAATPAAG	2040
Db	1981	GATTTTGTATTCATTATTTCCAGGCTCTGTATCTTAGAAATAGTGTGTTTCAAAATPAAG	2040
QY	2041	GGATCAATCTCATTTAGCAGGTAGTGAAGAACCCATGGCTGGGAAATGGAAGTGAAGCTGCCG	2100
Db	2041	GGATCAATCTCATTTAGCAGGTAGTGAAGAACCCATGGCTGGGAAATGGAAGTGAAGCTGCCG	2100
QY	2101	ACTGTGCACTGCTGCTGCAACGCTGTGAATGAGAGTCCAGTATTTCTTTCTTGACAG	2160
Db	2101	ACTGTGCACTGCTGCTGCAACGCTGTGAATGAGAGTCCAGTATTTCTTTCTTGACAG	2160
QY	2161	GACATCTCAACTCTTTTAAACATTAAAGACTCATTTGTGCTCTTGATCCAGCAGGCC	2220
Db	2161	GACATCTCAACTCTTTTAAACATTAAAGACTCATTTGTGCTCTTGATCCAGCAGGCC	2220
QY	2221	TTGATGCAATGGAAGGTATTATAGTCCCTGTGCTTTACAACCTTGACAGGACATGTGGT	2280
Db	2221	TTGATGCAATGGAAGGTATTATAGTCCCTGTGCTTTACAACCTTGACAGGACATGTGGT	2280
QY	2281	TATTTGCAAAATTTGACACTGACGCGGACCCCAAGAAATGTAATTAATTTGATAAACCATAAGG	2340
Db	2281	TATTTGCAAAATTTGACACTGACGCGGACCCCAAGAAATGTAATTAATTTGATAAACCATAAGG	2340

RESULT 3						PRI 13-DEC-2000
AF320293						
LOCUS	AF320293	2340 bp	mRNA	linear		
DEFINITION	Homo sapiens ABCG5 (ABCG5) mRNA,			complete cds.		
ACCESSION	AF320293					
VERSION	AF320293.1	GI:11692799				
KEYWORDS						
SOURCE	Homo sapiens (human)					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 2340)
Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultze, J.,

TITLE	Kwiterovich, P., Shan, B., Barnes, R. and Hobbs, H.H.
JOURNAL REFERENCE	Accumulation of Dietary Cholesterol in Stoseroolemia Caused by Mutations in Adjacent ABC Transporters Science (2001) In Press
AUTHORS	2 (bases 1 to 2340)
TITLE	Bergre, K.E., Tian, H., Graf, G.A., Yu, L., Grisham, N.V., Schultz, J.,
JOURNAL	Kwiterovich, P., Shan, B., Barnes, R. and Hobbs, H.H.
FEATURES	Direct Submission Submitted (09-NOV-2000) Molecular Genetics, University of Texas, Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9046, USA Location/Qualifiers

```

source
I. .2340
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .2340
/gene="ABCG5"
107. .2062
/gene="ABCG5"
/note="ATP-binding cassette, subfamily G, member 5"
/product="ABCG5"
/codon_start=1
/protein_id="AAC40003.1"
/db_xref="GI:11692800"
/translation="MGDLSLTTPGSGIGLVNVRGNSLEGAPATAPAEPSIGITAS
YSVSHRYRPMMDITSCROOQWTILKDVSLYVSGGOMICISGSGCKTILDMRSR
LGAGRTGIEVYNGRALRREORDEPSYLSQDPTLLSLTYETLHYTLAILRKN
PGSFQKRVKVAEELSLSHVADRLLIGNSLGGSTERRRNVSTAAQILQPKVWLDE
PTMGIDCMTRNQIIVLVLELARRNRILVLTIHQPRSELFLQFKIAILSGELIFGCT
PRLDLPFEDCGPCPEHSNPFEDYNDLTSVDQSKEREIETSRVOMTEASAKRSKI
CHTTLNIEEMKRLKTLIPWPEFTKQSPGVESKIGVLLRVETNLRVNTKLAIVTTLQ

```


Mon Jul, 28 09:43:27 2003

us-09-989-981a-5.rge

Page 5

N I N G L I G L F E F V L I R N S N T K G A I O D R V G L L I O Y V G A T P T T G M L N A V N L F E F V L R A V S D
 Q E O S D G L K O R K M R L A L A H L T P F S V A A T M I E S S V C T W T L G L E A D V A P F G F S A L L A
 P H L I G E L F I L L I G I O V N P N I V N S V A L L I S T A G L V O S G F L R N I O E D I P K I I S Y F P L
 F O K S E L I V N E R Y G L N F P C G S S N S V T N P M C A F O G I O F I E K T C P G A S R S F T N N E
 L L I S F P A L V I L G I V F E K I R D H I S R *
 541 a 601 c 598 g 600 t
 BASE COUNT
 ORIGIN

Query Match	100.0%	Score 2340;	DB 9;	Length 2340;
Best Local Similarity	100.0%;	Pred. N.O. 0;		
Matches 2340; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

QY	1	GTCAGTGGAGCAGGACAGGCGAGTCTGCCACGGGCTCCCAACTGAAGCCACTGTGGGA	60
Db	1	GTACAGGTGGAGCAGGCGAGGCGAGTCTGCCACGGGCTCCCAACTGAAGCCACTGTGGGA	60
QY	61	GGGTCCGGCCACAGAAAATTTGGCCCAAGCTTGGCTGTGGCCATGGGTACCTCTC	120
Db	61	GGGTCCGGCCACAGAAAATTTGGCCCAAGCTTGGCTGTGGCCATGGGTACCTCTC	120
QY	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAAGTAACAGAGGCTCCGAGACTCCCT	180
Db	121	ATCTTTGACCCCGGAGGGTCCATGGGTCTCCAAAGTAACAGAGGCTCCGAGACTCCCT	180
QY	181	GGAGGGGGCTCTCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCACTCTCATGCTCTTA	240
Db	181	GGAGGGGGCTCTCTGCCACCGCCCCGGAGCCTCACAGCCTGGGCACTCTCATGCTCTTA	240
QY	241	CAGCGTCAGCCACCGCGTGAAGGCGCTGGTGGGAAATCACATCTTTGCCGAGCAGTGGAC	300
Db	241	CAGCGTCAGCCACCGCGTGAAGGCGCTGGTGGGAAATCACATCTTTGCCGAGCAGTGGAC	300
QY	301	CAGCGAGATCTCAAAAGATGTCTCTTTGACGTGAGAGAGCGGCGACATCATGTGCTACTCT	360
Db	301	CAGCGAGATCTCAAAAGATGTCTCTTTGACGTGAGAGAGCGGCGACATCATGTGCTACTCT	360
QY	361	AGGAAAGCTAAGGCTCCGGGAAAACACACGCTGTGACAGCCATCTCCGGAGGCTGGGGCG	420
Db	361	AGGAAAGCTAAGGCTCCGGGAAAACACACGCTGTGACAGCCATCTCCGGAGGCTGGGGCG	420
QY	421	CGCGGGGACCTTCCTGGGGAGGTGTATGAAACGGCCGGGCGCTGTGCGCGGAGCAGTT	480
Db	421	CGCGGGGACCTTCCTGGGGAGGTGTATGAAACGGCCGGGCGCTGTGCGCGGAGCAGTT	480
QY	481	CCAGAGCTGCTTCTCTACGTCTGTGAGAGCAGCACCTCTGAGGACCTCACCCGAGCG	540
Db	481	CCAGAGCTGCTTCTCTACGTCTGTGAGAGCAGCACCTCTGAGGACCTCACCCGAGCG	540
QY	541	CGAGACGCTGCTTCTCTACGTCTGTGAGAGCAGCACCTCTGAGGACCTCACCCGAGCG	600
Db	541	CGAGACGCTGCTTCTCTACGTCTGTGAGAGCAGCACCTCTGAGGACCTCACCCGAGCG	600
QY	601	GAGAGAGGTGGAGGCGTGCATGGGAGGTGAGTGCAGCAATCTGGCAGACCTGAGT	660
Db	601	GAGAGAGGTGGAGGCGTGCATGGGAGGTGAGTGCAGCAATCTGGCAGACCTGAGT	660
QY	661	TGGCACTACACACTTGGGGGGCATTTTCCAGGGGTGAGCGCGCGGGTCTCCATCGCAGC	720
Db	661	TGGCACTACACACTTGGGGGGCATTTTCCAGGGGTGAGCGCGCGGGTCTCCATCGCAGC	720
QY	721	CCAGCTGCTCAGAGTTCCTAAGGTATGCTTTGATGAGCCAAACACAGCGCTGAGCTG	780
Db	721	CCAGCTGCTCAGAGTTCCTAAGGTATGCTTTGATGAGCCAAACACAGCGCTGAGCTG	780
QY	781	CATGACTGCTAATCATGATTTGTGCTCTCTGTTGAAACGAGTCCGAGAACCGAATTTGT	840
Db	781	CATGACTGCTAATCATGATTTGTGCTCTCTGTTGAAACGAGTCCGAGAACCGAATTTGT	840
QY	841	GGTTCTACCAATTCACAGCCCGGTTCTGAGAGCTTTTTCAGCTTTTGAACAAATTCGAT	900
Db	841	GGTTCTACCAATTCACAGCCCGGTTCTGAGAGCTTTTTCAGCTTTTGAACAAATTCGAT	900
QY	901	CCTAGAGTTCGAGAGCTGATTTTCTGTGGCACGCCAGCGGAAATGCTGATTTCTTCAA	960

Db	901	CGTAGCTTCGGAGAGCTGATTTTCTGTGCGACGGCAGCGGAAATGCTTGATTTCTTCAA	960
QY	961	TCAGTCGGGTTACCCCTGTCTCGAACATWCAACCCCTTTGACTCTCATATGAGACCTGAC	1020
Db	961	TCAGTCGGGTTACCCCTGTCTCGAACATWCAACCCCTTTGACTCTCATATGAGACCTGAC	1020
QY	1021	GTCACTGGCATACCCAAAGCAGGAAGCGGAAATAGAAACCTCCAGAGAGCTCCAGATGAT	1080
Db	1021	GTCACTGGCATACCCAAAGCAGGAAGCGGAAATAGAAACCTCCAGAGAGCTCCAGATGAT	1080
QY	1081	AGAAATCGCTCTCAGGAATAACGCAATTTGTCATPAAACTTTGAAAGATATTGAAAGAT	1140
Db	1081	AGAAATCGCTCTCAGGAATAACGCAATTTGTCATPAAACTTTGAAAGATATTGAAAGAT	1140
QY	1141	GAAGACCCGAAAGAGTTACCAATGTCTCTTCAAACCAAGATTCCTCGAGAGTTT	1200
Db	1141	GAAGACCCGAAAGAGTTACCAATGTCTCTTCAAACCAAGATTCCTCGAGAGTTT	1200
QY	1201	CTCTTAACTGGGTGTTCTCTGAGAGAGTACAAAGAACTTGGTGAGAAATTAAGCTGGC	1260
Db	1201	CTCTTAACTGGGTGTTCTCTGAGAGAGTACAAAGAACTTGGTGAGAAATTAAGCTGGC	1260
QY	1261	AGTGATTAACGGGTCTCTCTCAGAAATCTCATATGGGTTTGTTCCTCTTCTTCGTCT	1320
Db	1261	AGTGATTAACGGGTCTCTCTCAGAAATCTCATATGGGTTTGTTCCTCTTCTTCGTCT	1320
QY	1321	CGCGGTCCGGAAGCATGTGCTTAAAGGGGTGCTATCAGAGACCGGCTAGTCTCCCTTACA	1380
Db	1321	CGCGGTCCGGAAGCATGTGCTTAAAGGGGTGCTATCAGAGACCGGCTAGTCTCCCTTACA	1380
QY	1381	GTTTGTGGGGCCACACCCCTACACAGGCATCTGAAGCGTGTGAATCTGTTTCCCGAGCT	1440
Db	1381	GTTTGTGGGGCCACACCCCTACACAGGCATCTGAAGCGTGTGAATCTGTTTCCCGAGCT	1440
QY	1441	CGGAGCTTCAAGCCAGCAGAGAGCTCAGAGAGCGGCTCTACAGAAAGTGGCAGATGATGCT	1500
Db	1441	CGGAGCTTCAAGCCAGCAGAGAGCTCAGAGAGCGGCTCTACAGAAAGTGGCAGATGATGCT	1500
QY	1501	GGCCATGACATGACAGCTCCGCCCTTCAGAGTGTGGCCACCATGATTTTCAGACAGTGT	1560
Db	1501	GGCCATGACATGACAGCTCCGCCCTTCAGAGTGTGGCCACCATGATTTTCAGACAGTGT	1560
QY	1561	GTGCTACTGAGAGCTGTGGGCTTACATTCCTGAGAGTGGCCGATTTGGGATATTTTCTGTGC	1620
Db	1561	GTGCTACTGAGAGCTGTGGGCTTACATTCCTGAGAGTGGCCGATTTGGGATATTTTCTGTGC	1620
QY	1621	TCCTGTGGCCCCCACCATTAAATGGTGAAATTTCTAACCTCTGTGCTACTTGGTATGCGCA	1680
Db	1621	TCCTGTGGCCCCCACCATTAAATGGTGAAATTTCTAACCTCTGTGCTACTTGGTATGCGCA	1680
QY	1681	AAATCCAAATATATCAACAGTGTATGAGGCTCTGTCTGTCCATTTGGCGGGGCTGTTGTGG	1740
Db	1681	AAATCCAAATATATCAACAGTGTATGAGGCTCTGTCTGTCCATTTGGCGGGGCTGTTGTGG	1740
QY	1741	ATCTGGAATTCCTCAGAAACATACAGAAATGCCCATTTCTTTAAATCATACGATATTT	1800
Db	1741	ATCTGGAATTCCTCAGAAACATACAGAAATGCCCATTTCTTTAAATCATACGATATTT	1800
QY	1801	TACATTTCCAAAATATTTGACAGTGAATCTTGTAGTCAATGAGATTTCTACGAGACTGAATTT	1860
Db	1801	TACATTTCCAAAATATTTGACAGTGAATCTTGTAGTCAATGAGATTTCTACGAGACTGAATTT	1860
QY	1861	CACCTGTGGCAGCTCAAAATTTTCTGTGACACTAAATCCAAAGTGTGCCTTACCTCAAGG	1920
Db	1861	CACCTGTGGCAGCTCAAAATTTTCTGTGACACTAAATCCAAAGTGTGCCTTACCTCAAGG	1920
QY	1921	AATTCATTTAATGAGAAAACGCGCCAGGTGCAAAATTAAGATTTACAAATGACTTTCT	1980
Db	1921	AATTCATTTAATGAGAAAACGCGCCAGGTGCAAAATTAAGATTTACAAATGACTTTCT	1980
QY	1981	GATTTTGTATTCATTTATTCACGCTGTGCACTCCAGGAATAGTGTTCCTCAAAATAG	2040

Db	1991	GATTTTGTAATTCATTATTATTCACAGCTCTTGATCCATCCPAGAAATAGTGTGTTTTCAAAATPAG	2040	
Oy	2041	GGATCATTCTCAATTTACGAGTAGTAGTAAGAACCCATGGCTGGGAAATAGAGTAGACGTGCC	2100	
Db	2041	GGATCATTCTCAATTTACGAGTAGTAGTAAGAACCCATGGCTGGGAAATAGAGTAGACGTGCC	2100	
Oy	2101	ACTGTGACGACCTGCTTGAACGCTGTGAATAGAGTAGTGCATGTATTTCTTTCTTGTGACAG	2160	
Db	2101	ACTGTGACGACCTGCTTGAACGCTGTGAATAGAGTAGTGCATGTATTTCTTTCTTGTGACAG	2160	
Oy	2161	GACATCTCAAGCTTTTATACCATTTAAGACATCTCAATTTGCTCTGTGATCCACAGAGCC	2220	
Db	2161	GACATCTCAAGCTTTTATACCATTTAAGACATCTCAATTTGCTCTGTGATCCACAGAGCC	2220	
Oy	2221	TTGATGCAATGAGAGGCTTTATAGTCCCTTGTCTTTACAATTCGACAGGACATGTGCT	2280	
Db	2221	TTGATGCAATGAGAGGCTTTATAGTCCCTTGTCTTTACAATTCGACAGGACATGTGCT	2280	
Oy	2281	TATTTGGAAATTTGACTGACGACGACCCAGACAAATGTAAATATTCATATAAAGCTATGCG	2340	
Db	2281	TATTTGGAAATTTGACTGACGACGACCCAGACAAATGTAAATATTCATATAAAGCTATGCG	2340	
RESULT 4	AF312715	2740 bp	linear	PRI 14-JUN-2001
LOCUS	Homo sapiens sterolin (ABCG5) mRNA, complete cds.			
ACCESSION	AF312715			
VERSION	AF312715.2	GI:14423628		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
AUTHORS	Lee,M.H., Lu,K.R., Hazard,S., Yu,H., Shulenn,S., Hidaka,H., Solen,G., Dean,M. and Patel,S.B.			
TITLE	Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption			
JOURNAL	Nat. Genet. 27 (1), 79-83 (2001)			
MEDLINE	20578753			
PUBMED	11138003			
REFERENCE	2 (bases 1 to 2740)			
AUTHORS	Lu,K., Lee,M.-H. and Patel,S.B.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St. S7B541, Charleston, SC 29403, USA			
COMMENT	On Jun 14, 2001 this sequence version replaced gi:12382303.			
FEATURES	Location/Qualifiers			
source	1..2740			

N,INIGLIFLFEVLIVRSDSVNLKGAIOADVGLLYQGVGATPTTGM,NAVNLEPVLRAVSD
 QESDGLXQKQNMALALAHVLPFSVAHMLIESVCYVTLGLBHNARFGYFSALALA
 PHILGELITVLGLIYONPNIIVNSVAHMLIESVGLYVSGFLRNLIQEPKILSYFT
 FQKSCFLVIVNEFYGNLTGSSNSVSTWPMKCAFTQGIQFIETKCPGASRFTNPF
 LILSPFALVILGLIIVFKIRIDHLSIR*
 669 a 670 c 702 g 699 t

Query Match	Similarity	100.0%	Score 2340;	DB 9;	Length 2740;
Best Local	Similarity	100.0%	Pred. No. 0;	Mismatches 2340;	Conservative 0;
QY	1	GTGAGTGGAGCAGCAGGAGGAGTGTGGCAGCGGGTCCCAACTGAACCACTGAGGGA	60		
Db	35	GTGAGTGGAGCAGCAGGAGGAGTGTGGCAGCGGGTCCCAACTGAACCACTGAGGGA	94		
QY	61	GGGTCCGGCCACCAAGAAATTTGGCCAGCTTGTGCTGTGGCCATGGGTGACCTTC	120		
Db	95	GGGTCCGGCCACCAAGAAATTTGGCCAGCTTGTGCTGTGGCCATGGGTGACCTTC	154		
QY	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAACTAAACAGAGGCTCCAGAGCTCCT	180		
Db	155	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAACTAAACAGAGGCTCCAGAGCTCCT	214		
QY	181	GGAGGGGGGTCTCTGGCCACCGCCCGGAGCCTCACAGCCTGGGCACTCCCAATGGCTCTA	240		
Db	215	GGAGGGGGGTCTCTGGCCACCGCCCGGAGCCTCACAGCCTGGGCACTCTCCATGGCTCTA	274		
QY	241	CAGGCTCAGCCACCGCGTGAAGGCCCTGGTGGGACATCATCTTGGCCGGCAGAGTGGAC	300		
Db	275	CAGGCTCAGCCACCGCGTGAAGGCCCTGGTGGGACATCATCTTGGCCGGCAGAGTGGAC	334		
QY	301	CAGGAGATTCCTCAAGATGTCTCTTGTAGTGGAGAGCGGGCAGATCATGTCACTCT	360		
Db	335	CAGGAGATTCCTCAAGATGTCTCTTGTAGTGGAGAGCGGGCAGATCATGTCACTCT	394		
QY	361	AGGAAGCTCAGAGCTCCGGGAAACACACGCTCTGGAGCCATGTCCGGAGAGGCTGGGGCG	420		
Db	395	AGGAAGCTCAGAGCTCCGGGAAACACACGCTCTGGAGCCATGTCCGGAGAGGCTGGGGCG	454		
QY	421	CGCGGGGACCTTCTCGGGGGAGGTGTATGTAAACGGCCGGCGCTGGCGGGAGCAGTT	480		
Db	455	CGCGGGGACCTTCTCGGGGGAGGTGTATGTAAACGGCCGGCGCTGGCGGGAGCAGTT	514		
QY	481	CCAGAGCTGTCTCTCTACGTCGTCAGAGAGGACACCTGTGTGAGAGGCTCAACCGGCG	540		
Db	515	CCAGAGCTGTCTCTCTACGTCGTCAGAGAGGACACCTGTGTGAGAGGCTCAACCGGCG	574		
QY	541	CGAGAGCTGTACACACCGCGCTGCGGCCATCGCGCGGGCAGTCCCGGCTCTTCCA	600		
Db	575	CGAGAGCTGTACACACCGCGCTGCGGCCATCGCGCGGGCAGTCCCGGCTCTTCCA	634		
QY	601	GAAAGAGTGGAGGCCCTGTCATGGCAGAGCTGATGTATGAGCCATGTGGCAGACGATGAT	660		
Db	635	GAAAGAGTGGAGGCCCTGTCATGGCAGAGCTGATGTATGAGCCATGTGGCAGACGATGAT	694		
QY	661	TGGCAACATCAGCTTGGGGGGCATTTCACAGGGGAGCGGGCGCGGGGTCCCAATCGCAGC	720		
Db	695	TGGCAACATCAGCTTGGGGGGCATTTCACAGGGGAGCGGGCGCGGGGTCCCAATCGCAGC	754		
QY	721	CCAGCTGTCCAGAGATCCTAAGGTCATGCTTTGTGATGAGCAACACACAGAGCCTGAGCTG	780		
Db	755	CCAGCTGTCCAGAGATCCTAAGGTCATGCTTTGTGATGAGCAACACACAGAGCCTGAGCTG	814		
QY	781	CATACAGCTAATCAGATGTGCTGCTCTGGGTGGAACTGGCTGCGAGAACGGAATGTG	840		
Db	815	CATACAGCTAATCAGATGTGCTGCTCTGGGTGGAACTGGCTGCGAGAACGGAATGTG	874		
QY	841	GGTTCACCATTCACACAGCCCGTGTCTGAGCTTTTTCAGCTCTTTTGAACAAATTTGCAAT	900		
Db	875	GGTTCACCATTCACACAGCCCGTGTCTGAGCTTTTTCAGCTCTTTTGAACAAATTTGCAAT	934		
QY	901	CTGAGACTTGGAGAGCTGATTTTCTGTGGCAGCAGCGGAATGCTGATTTTCTTCA	960		

```

|||||
935 CTTGAGCTTGGGAGAGCTGATTTTCTGTGACGCCCGACGCGAAATGCTGATTTCTTCA 994
OY 961 TGAAGCTGTTTACCTTGTCTCTGACATCAAAACCCCTTTTACTTCTATATGAGACCTGAC 1020
Db 995 TGAGCTGGGTTTACCTTGTCTCTGACATCAAAACCCCTTTTACTTCTATATGAGACCTGAC 1054
OY 1021 GTCAGTGGATACCCAAAGCAAGGAAATAGAAACCTCCAGAGAGTCCAGATGAT 1080
Db 1055 GTCAGTGGATACCCAAAGCAAGGAAATAGAAACCTCCAGAGAGTCCAGATGAT 1114
OY 1081 AGAATCTGCTTACAGAAATACAGCAATTTGTCTATAAAACCTTGAAGAAATTTGAAAGAAAT 1140
Db 1115 AGAATCTGCTTACAGAAATACAGCAATTTGTCTATAAAACCTTGAAGAAATTTGAAAGAAAT 1174
OY 1141 GAAAGACCTGAAACGTTTACCAATGTTCTCTTCAAAACCAAGATTTCTCTGAGATTTT 1200
Db 1175 GAAAGACCTGAAACGTTTACCAATGTTCTCTTCAAAACCAAGATTTCTCTGAGATTTT 1234
OY 1201 CTCTAAACTGGGTGTTCTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATAGAGTGGC 1260
Db 1235 CTCTAAACTGGGTGTTCTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATAGAGTGGC 1294
OY 1261 AGTGAATTAAGCGCTCTCTCTCTGAGAAATGATGATGGTTTGTCTCTCTCTCTCTCTCTCT 1320
Db 1295 AGTGAATTAAGCGCTCTCTCTCTCTGAGAAATGATGATGGTTTGTCTCTCTCTCTCTCTCTCT 1354
OY 1321 GGGGGTCCGGAAGAAATGCTTAAGAGTGTATATCCAGAGACCGGGTGTCTCTCTTACCA 1380
Db 1355 GCGGGGTCCGGAAGAAATGCTTAAGAGTGTATATCCAGAGACCGGGTGTCTCTCTTACCA 1414
OY 1381 GTTGTGGGGCCACCCCGCTTACACAGGCAATGCTGAACGCGTGAATCTGTTTCCCGGTCT 1440
Db 1415 GTTGTGGGGCCACCCCGCTTACACAGGCAATGCTGAACGCGTGAATCTGTTTCCCGGTCT 1474
OY 1441 GCGAGCTGTACGAGCAGCAGAGAGTGAAGAGCGGCTCTACACAGAGTGCAGATGATGCT 1500
Db 1475 GCGAGCTGTACGAGCAGCAGAGAGTGAAGAGCGGCTCTACACAGAGTGCAGATGATGCT 1534
OY 1501 GGGCTATGACATGACAGCTGCTCCCTCCCTTACAGCGTGTGTGTCACACATATTTTCCAGCTGT 1560
Db 1535 GGGCTATGACATGACAGCTGCTCCCTCCCTTACAGCGTGTGTGTCACACATATTTTCCAGCTGT 1594
OY 1561 GTGCTACTGAGACGCTGGGCTTACATCTGAGTGTGCGCGATTTTGATATTTTCTCTGCG 1620
Db 1595 GTGCTACTGAGACGCTGGGCTTACATCTGAGTGTGCGCGATTTTGATATTTTCTCTGCG 1654
OY 1621 TCTCTTGGCCCCCAGCTTAATGTTGTAATTTCTTACTTCTTGTGCTACTTGTATCTGCCA 1680
Db 1655 TCTCTTGGCCCCCAGCTTAATGTTGTAATTTCTTACTTCTTGTGCTACTTGTATCTGCCA 1714
OY 1681 AATATCAAAATATATGATCAAGAGTGTAGTGTGCTGCTGCTCATTTGGCGGGGTGCTTGTGG 1740
Db 1715 AATATCAAAATATATGATCAAGAGTGTAGTGTGCTGCTGCTCATTTGGCGGGGTGCTTGTGG 1774
OY 1741 ATCTGATTTCTCTAGAAACATACAAAGAAATGCGCATTTCTTTTAAATATACATGATTTT 1800
Db 1775 ATCTGATTTCTCTAGAAACATACAAAGAAATGCGCATTTCTTTTAAATATACATGATTTT 1834
OY 1801 TACATTTCAAAAATATATGAGTGTGATTTCTTGTAGTCAATGATTTTACGAGTCAATTT 1860
Db 1835 TACATTTCAAAAATATATGAGTGTGATTTCTTGTAGTCAATGATTTTACGAGTCAATTT 1894
OY 1861 CACTTGTGCGAGCTCAAAATGTTTGTGTGCAACTAATTCATATGTTGCTTCACTCAAG 1920
Db 1895 CACTTGTGCGAGCTCAAAATGTTTGTGTGCAACTAATTCATATGTTGCTTCACTCAAG 1954
OY 1921 AATTCATTTCAATTAAGAAACCTGGCGAGGTGCAACATCTAGATTCACAAATATCTTCT 1980
Db 1955 AATTCATTTCAATTAAGAAACCTGGCGAGGTGCAACATCTAGATTCACAAATATCTTCT 2014
OY 1981 GATTTTGTATTCATTTATTCAGCTCTTGTCTATCTCTAGAAATAGTTGTTTCAAAATAG 2040
|||||

```

```

Db 2015 GATTTTGTATTCATTTATTCACAGCTCTTGTGATCTCTAGAAATAGTTGTTTCAAAATAG 2074
OY 2041 GGATCATCTCTATTAAGCAGGTAGTGAAGCCATGGCTGGGAAATGGAAGTGAAGCTGCCG 2100
Db 2075 GGATCATCTCTATTAAGCAGGTAGTGAAGCCATGGCTGGGAAATGGAAGTGAAGCTGCCG 2134
OY 2101 ACTGTGATGACAGCTCTCTGAAAGCTGAAATGAGAGGCAATATTTCTTCTTGTACAG 2160
Db 2135 ACTGTGATGACAGCTCTCTGAAAGCTGAAATGAGAGGCAATATTTCTTCTTGTACAG 2194
OY 2161 GACATCTCAAGCTCTTTTAAACATTAAGACTCAATTTGTGCTCTTGTGATCCAGAGGCC 2220
Db 2195 GACATCTCAAGCTCTTTTAAACATTAAGACTCAATTTGTGCTCTTGTGATCCAGAGGCC 2254
OY 2221 TTGAATGCAATGGAAGTGTATTAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2280
Db 2255 TTGAATGCAATGGAAGTGTATTAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2314
OY 2281 TATTTGGAATTTGACTGAGCGGACCCAGCAAGATGTAATATATTCATTAACCTATGG 2340
Db 2315 TATTTGGAATTTGACTGAGCGGACCCAGCAAGATGTAATATATTCATTAACCTATGG 2374

RESULT 5
AX456520
LOCUS AX456520 2516 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 42 from Patent WO0227016.
ACCESSION AX456520
VERSION AX456520.1 GI:21715410
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Patel,S.B. and Dean,M.
TITLE Gene involved in dietary sterol absorption and excretion and uses
JOURNAL Patent: WO 0227016-A 42 04-APR-2002;
THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,
Shallendra B. (US) ; Dean, Michael (US)
FEATURES
source 1..2516
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/label="Primer"
BASE COUNT 601 a 631 c 636 g 648 t
ORIGIN
Query Match 99.9%; Score 2338.4; DB 6; Length 2516;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GTCAGTGGAGAGAGGAGGAGGAGTGTGCGACGGGCTCCCACTGAAGCACTTGGGA 60
35 GTCAGTGGAGAGAGGAGGAGGAGTGTGCGACGGGCTCCCACTGAAGCACTTGGGA 94
61 GGTCCGGGACACAGAAATTTGGCCAGCTTGTGCTGCTGTTGGCCATGGGTGACCTTC 120
95 GGTCCGGGACACAGAAATTTGGCCAGCTTGTGCTGCTGTTGGCCATGGGTGACCTTC 154
121 ATCTTGAACCCCGGAGGCTCATGGGTCTCAAGTAAACAGAGGCTCCAGAGCTTCCCT 180
155 ATCTTGAACCCCGGAGGCTCATGGGTCTCAAGTAAACAGAGGCTCCAGAGCTTCCCT 214
181 GGAGGGGCTCTCTGACAGCGCCCGGAGGCTCAAGCTGGGCAATCTCTCATGCTCTTA 240
215 GGAGGGGCTCTCTGACAGCGCCCGGAGGCTCAAGCTGGGCAATCTCTCATGCTCTTA 274
241 CAGCTGACACACCGGTGAGGCTGTGGGAGCATCATCTTGGCGGAGCAGCTGGAG 300
275 CAGCTGACACACCGGTGAGGCTGTGGGAGCATCATCTTGGCGGAGCAGCTGGAG 334

```

OY 301 CAGCAGATCTCTAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGCACTCT 360
|||||
DB 335 CAGGCAAGATCTCTAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGCACTCT 394
|||||
OY 361 AGGAAGCTCAGGCTCCGGGAAAAACAACGCTGCTGAGAGCGGCGAGATCATGTGCACTCT 420
|||||
DB 395 AGGAAGCTCAGGCTCCGGGAAAAACAACGCTGCTGAGAGCGGCGAGATCATGTGCACTCT 454
|||||
OY 421 CCGGGGAGACCTTCTCGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCGGAGAGATT 480
|||||
DB 455 CCGGGGAGACCTTCTCGGGGAGGTGTATGTGAACGGCCGGGCGCTGCGCGGAGAGATT 514
|||||
OY 481 CCAGGACTGCTTCTCTAGCTCTCTGAGAGCGACACCTGCTGAGAGAGCTTCAACGCTGCG 540
|||||
DB 515 CCAGGACTGCTTCTCTAGCTCTCTGAGAGCGACACCTGCTGAGAGAGCTTCAACGCTGCG 574
|||||
OY 541 CGAGAGCTGCACTACACCGCGCTGCTGGCCATCCCGCGGCAATCCCGGCTCTTCCA 600
|||||
DB 575 CGAGAGCTGCACTACACCGCGCTGCTGGCCATCCCGCGGCAATCCCGGCTCTTCCA 634
|||||
OY 601 GAAGAAGTGAAGGCGGCTGATGAGAGAGCTGAGCTGAGCATGTGCGAGACGAGCTGAT 660
|||||
DB 635 GAAGAAGTGAAGGCGGCTGATGAGAGAGCTGAGCTGAGCATGTGCGAGACGAGCTGAT 694
|||||
OY 661 TGGCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGCGCGGGTCTCCATCCGACG 720
|||||
DB 695 TGGCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGCGCGGGTCTCCATCCGACG 754
|||||
OY 721 CCAGGCTGCGAGAGATCTTAAGTCAATGCTGTTTGTATGAGCCACACAGGCTGACATG 780
|||||
DB 755 CCAGGCTGCGAGAGATCTTAAGTCAATGCTGTTTGTATGAGCCACACAGGCTGACATG 814
|||||
OY 781 CATGACTGCTTAATCAGATTTGTGCTCTCTGAGTGAACGAGCTCGAGAGACCGAATTTG 840
|||||
DB 815 CATGACTGCTTAATCAGATTTGTGCTCTCTGAGTGAACGAGCTCGAGAGACCGAATTTG 874
|||||
OY 841 GGTCTCTACCAATTCACACGCCCCGTTCTGAGCTTTTTCAGCTTTTTCAGAAAATTTCCAT 900
|||||
DB 875 GGTCTCTACCAATTCACACGCCCCGTTCTGAGCTTTTTCAGCTTTTTCAGAAAATTTCCAT 934
|||||
OY 901 CCGGAGCTTCGGGAGAGCTATTTTCTGCGGACGCCAGCGGGAATGCTGATTTCTTCAA 960
|||||
DB 935 CCGGAGCTTCGGGAGAGCTATTTTCTGCGGACGCCAGCGGGAATGCTGATTTCTTCAA 994
|||||
OY 961 TGAAGTCCGCTTACCTTGTCTGTAACATTTCAAAACCTTTTGAATTTATATGAGACTGAC 1020
|||||
DB 995 TGAAGTCCGCTTACCTTGTCTGTAACATTTCAAAACCTTTTGAATTTATATGAGACTGAC 1054
|||||
OY 1021 GTCAAGTGAATCCCAAGAGAGGAAATGAAATTCGCAAGAGAGTCCAGATGAT 1080
|||||
DB 1055 GTCAAGTGAATCCCAAGAGAGGAAATGAAATTCGCAAGAGAGTCCAGATGAT 1114
|||||
OY 1081 AGAATCTGCTCAAGAAATCAGCAATTTGTATATAAATTTGAAAGATTTGAAAGAT 1140
|||||
DB 1115 AGAATCTGCTCAAGAAATCAGCAATTTGTATATAAATTTGAAAGATTTGAAAGAT 1174
|||||
OY 1141 GAAGACCTGAAAGAGTTTACCAATGTTCTTTCAAAACCAAGATTTCTCGAGATTTT 1200
|||||
DB 1175 GAAGACCTGAAAGAGTTTACCAATGTTCTTTCAAAACCAAGATTTCTCGAGATTTT 1234
|||||
OY 1201 CTCTAACTGGGTGTTCTCTGAGAGAGTGAAGAAATTTGTTGAGAAATTAAGCTGAC 1260
|||||
DB 1235 CTCTAACTGGGTGTTCTCTGAGAGAGTGAAGAAATTTGTTGAGAAATTAAGCTGAC 1294
|||||
OY 1261 AGTGAATTAAGGCTTCCCTTCAAGATCTGATGTTGTTGCTTCTCTGCTTCT 1320
|||||
DB 1295 AGTGAATTAAGGCTTCCCTTCAAGATCTGATGTTGTTGCTTCTCTGCTTCT 1354
|||||
OY 1321 GCGGGTCCGAAACCAATGTGCTTAAAGGTGCTATCCAGAGACCGGATGCTCTCTTACCA 1380
|||||
DB 1355 GCGGGTCCGAAACCAATGTGCTTAAAGGTGCTATCCAGAGACCGGATGCTCTCTTACCA 1414
|||||
OY 1381 GTTTGTGGGGCCGACCCGTTACACAGGAGCTGTAACGCTGTGAATCTGTCTCCGCTGCT 1440
|||||

DB 1415 GTTTGTGGGGCCGACCCGCTTACACAGGCAATGGTGAACGGTGTGAATCTGTTCCCTGCT 1474
|||||
OY 1441 GCGAGCTGTACGAGGACACAGAGAGTGAAGAGCGGCTTCAACAGAACTGCGACATGTGCT 1500
|||||
DB 1475 GCGAGCTGTACGAGGACACAGAGAGTGAAGAGCGGCTTCAACAGAACTGCGACATGTGCT 1534
|||||
OY 1501 GGGCTATGCACTCAGCTGCTCCCTTCAAGCGGTTGTTGCGACCATATTTTTCAGAGCTG 1560
|||||
DB 1535 GGGCTATGCACTCAGCTGCTCCCTTCAAGCGGTTGTTGCGACCATATTTTTCAGAGCTG 1594
|||||
OY 1561 GTGCTACTGAGAGCTGGGCTTACATTCCTGAGGTTGGCCGATTTGATTTTCTGCTGCG 1620
|||||
DB 1595 GTGCTACTGAGAGCTGGGCTTACATTCCTGAGGTTGGCCGATTTGATTTTCTGCTGCG 1654
|||||
OY 1621 TCTCTTGGCCCCCACTTAAATTTGTAATTTCTAACTCTTGTGCTAATTTGTAATGCTTCA 1680
|||||
DB 1655 TCTCTTGGCCCCCACTTAAATTTGTAATTTCTAACTCTTGTGCTAATTTGTAATGCTTCA 1714
|||||
OY 1681 AATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
|||||
DB 1715 AATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1774
|||||
OY 1741 ATCTGATTCCTCAGAAACATACAGAAATGCCCCATTTCTTTTAAATCATCAGTATTT 1800
|||||
DB 1775 ATCTGATTCCTCAGAAACATACAGAAATGCCCCATTTCTTTTAAATCATCAGTATTT 1834
|||||
OY 1801 TACATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
|||||
DB 1835 TACATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1894
|||||
OY 1861 CACTGTGGCAGCTGCAAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
|||||
DB 1895 CACTGTGGCAGCTGCAAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1954
|||||
OY 1921 AATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
|||||
DB 1955 AATTCGAAATTAATGTAACAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2014
|||||
OY 1981 GATTTTGTATGTAATTTATTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
|||||
DB 2015 GATTTTGTATGTAATTTATTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2074
|||||
OY 2041 GATTCATTCATTAAGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
|||||
DB 2075 GATTCATTCATTAAGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2134
|||||
OY 2101 ACTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
|||||
DB 2135 ACTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2194
|||||
OY 2161 GATTCATTCATTAAGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
|||||
DB 2195 GATTCATTCATTAAGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2254
|||||
OY 2221 TTGAATGCAATGGAAGTGGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
|||||
DB 2255 TTGAATGCAATGGAAGTGGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2314
|||||
OY 2281 TATTTGGAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTG 2340
|||||
DB 2315 TATTTGGAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTG 2374
|||||

RESULT 6
AX456519 1920 bp DNA linear PAT 06-JUL-2002
LOCUS AX456519
DEFINITION Sequence 41 from Patent WO0227016.
ACCESSION AX456519
VERSION AX456519.1 GI:21715409
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct

artificial sequences.

REFERENCE	AUTHORS	TITLE
1	Patel, S. B. and Dean, M.	Gene involved in dietary sterol absorption and excretion and uses therefor

BASE COUNT	440 a	503 c	486 g	491 t
ORIGIN				

Query Match	82.1%;	Score 1920;	DB 6;	Length 1920;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1920;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY	143	ATGGGCTCTCAAGTAAACAGAGGGGCTCCAGAGCTCCCTGGAGGGGGCTCTCTCCACGGCC	202
Db	1	ATGGGCTCTCAAGTAAACAGAGGGCTCCAGAGCTCCCTGGAGGGGGCTCTCTCCACGGCC	60
QY	203	CCGAGGCTCACAGGCTTGGGCAATCTCTCCATGGCTCTCTACAGGCTTAGCCACCGGCTGAGG	262
Db	61	CCGAGGCTCACAGGCTTGGGCAATCTCTCCATGGCTCTCTACAGGCTTAGCCACCGGCTGAGG	120
QY	263	CCCTGGTGGGACATCACATCTTTCGCCGCGACAGTGGACCGAGCAGATCCTCAAAAGATGTC	322
Db	121	CCCTGGTGGGACATCACATCTTTCGCCGCGAGAGTGGACCGAGCAGATCCTCAAAAGATGTC	180
QY	323	TCTCTGTACGTGAGAGAGCGGGGACATCTATGTCATCTTGGCATCTTAAAGAACTCAGGCTCGGGAAA	382
Db	181	TCTCTGTATGTGAGAGAGCGGGGACATCTATGTCATCTTAAAGAACTCAGGCTCGGGAAA	240
QY	383	ACCAAGCTGTGAGACCGCCATGTCCTGGAGAGCTGGGGCGCGCGGGGACCTTCTGAGGGAG	442
Db	241	ACCAAGCTGTGAGACCGCCATGTCCTGGAGAGCTGGGGCGCGCGGGGACCTTCTGAGGGAG	300
QY	443	GTTGATGTGAACGGGCGGGGCGCTGCGCGGGAGCAGTTCAGAGACTGCTTCTCTACGTC	502
Db	301	GTTGATGTGAACGGGCGGGGCGCTGCGCGGGAGCAGTTCAGAGACTGCTTCTCTACGTC	360
QY	503	CTGCAGAGACGACACCTCTGTGAGAGAGCTCAACGCTGGCGGAGAGCTGCTACACCGCG	562
Db	361	CTGCAGAGACGACACCTCTGTGAGAGAGCTCAACGCTGGCGGAGAGCTGCTACACCGCG	420
QY	563	CTGCTGGCCATCCGCGCGGCAATCCCGGCTCTCTTCAGAAAGAGTGGAGGCGCTCATG	622
Db	421	CTGCTGGCCATCCGCGCGGCAATCCCGGCTCTCTTCAGAAAGAGTGGAGGCGCTCATG	480
QY	623	GCAGAGTGTGCTGAGACCATGTGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGCG	682
Db	481	GCAGAGTGTGAGCTGAGACCATGTGTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGCG	540
QY	683	ATTATCCACGGGTGAGCGGCGCGGGGTCTCATGCGACACCCAGCTGCTCCAGAGATCCTTAAG	742
Db	541	ATTATCCACGGGTGAGCGGCGCGGGGTCTCATATCCAGACCCAGCTGCTCCAGAGATCCTTAAG	600
QY	743	GTCATGCTGTTTGTAGTGAAGCCACACACAGGCGCTGACTGCATGACTGCTATACAGATTGTC	802
Db	601	GTCATGCTGTTTGTAGTGAAGCCACACACAGGCGCTGACTGCATGACTGCTATACAGATTGTC	660
QY	803	GTCCTCTCTGTTGGAATCTGGCTCGCAGGAACCGAATTGTGGTCTCACACATTCACACGCC	862
Db	661	GTCCTCTCTGTTGGAATCTGGCTCGCAGGAACCGAATTGTGGTCTCACACATTCACACGCC	720
QY	863	CGTTCGAGGTTTTTTCAGCTCTTTCGACAAAATTGGCCATCTGAGCTTTCGGAGAGCTGATT	922
Db	721	CGTTCGAGGTTTTTTCAGCTCTTTCGACAAAATTGGCCATCTGAGCTTTCGGAGAGCTGATT	780

[illegible]

[illegible]

AUTHORS		TITLE		JOURNAL		FEATURES		BASE COUNT		ORIGIN		Query Match		Best Local Similarity		Matches 1664;		Conservative		0;		Mismatches 403;		Indels		3;		Gaps		1	
Gatel, S. B. and Dean, M.		Gene involved in dietary sterol absorption and excretion and uses thereof		Patent: WO 0227016-A 46 04-APR-2002; THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US); Shallendra B. (US); Dean, Michael (US)		Location/Qualifiers 1.2354 /organism="synthetic construct" /mol_type="genomic DNA" /db_xref="taxon:32630" /note="Primer"		573 a 604 c 594 g 583 t				60.2%; Score 1409.2; DB 6; Length 2354;		80.4%; Pred. No. 36-296;																	
QY	25	CTGCCAGGGGCTCCCACTGAGCCACTGTGGGAGGGGTCCGGCCACAGAAATTTGC	84	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	57	CTCCACTTGGCTCTCAGTTAAAGCTGCCCTGTGAGCCGAGCCAGGACCTAGAAATTCAC	116	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	85	CCAGCTTGTGCTGCTGGTGGCCATGGGTGACTCTCATCTTTTGACCCCGGAGGTCAT	144	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	117	TTTGCACTTGTCTGCTGCTAGCCATGGGTGAGCTGCCCTTCTGACCTCCAGAGGAGCCAG	176	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	145	GGGTCTCCAAAGTAAACAGAGGCTCCACAGGCTCCCTGTGAGGGGGCTCCGCCACCGGCC	204	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	177	AGGGGCTCACATCAACAGAGGCTCTGAGCTCCCTGTGAGGACAGGTTGGTTCACGGGCAC	236	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	205	GGAGCCT--CACAGCTGGGCGCATCTCCATGCTCTCAACAGCTCAGCCACCGCGTAG	261	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	237	AGAGGCTCGGCGACACTTAGGTGTCTCATGTGTCTCAACAGCTCAGCAACCGTGTGG	296	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	262	GCCGTGGGGGCAATCATATCTTGGCGGACAGAGTGAGACAGGCAATCTCAAGATGT	321	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	297	GCTTGGTGGGAACATCAATCATCTGACACAGAGGAGGGAAGGCAAACTCTCAAGATGT	356	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	322	CTCTTTGACGTGAGAGCGGGCAGATCATGTGCATCTCTAGAGAGCTCAGGCTCCGGA	381	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
DB	357	CTCCTTGATCATCAGAGAGTGGCGAGATTAATGTGATCTTTAGGACAGCTCAGGCTCAG	416	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
QY	382	AACCAAGCTGTGAGCGGCATGTCCGGAGG																													

Db 837 CCTCTCTGTGAGCTGGCTGCGAGGAGATTCGATTTGATTCACATCCACGACC 896
 QY 862 CCGTTCAGCTTTTTCAGCTCTTTTGACAAATTCACCTTGAGCTTGAGAGATGAT 921
 Db 897 TCCTCTAGCTCTTTCACACACTTCGACAAATTCGATCTGATTCAGAGAGTGT 956
 QY 922 TTTCTGTGACGCGCAGCGAATGCTTATTTCTCAATGACTGCGGTAACTCTTTC 981
 Db 957 GTTCTGTGAGACCCGACAGAGAGATCTGCTCTTCAATACAGCTGTATACCCCTG 1016
 QY 982 TGAACATTCAAACCTTTGACTTCTATATGAGACCTGACGATGATACCCAAACAA 1041
 Db 1017 TGAACATTCAAACCTTTGACTTCTATATGAGACCTGACGATGATACCCAAACAA 1076
 QY 1042 GGAACGGGAATGAAACCTTCAGAGAGCTCAGATGATGATGATGATGATGATG 1101
 Db 1077 AGAGCGGGAATGAAACCTTCAGAGAGCTCAGATGATGATGATGATGATGATG 1136
 QY 1102 AGCAATTTGCTATTAACCTTTGAGAGATTTGAAAGATGAAACCTGAAACCTTAC 1161
 Db 1137 TGAATCTATCAACAAATTCGAGAGATTTGAAAGATGAAACCTGAAACCTTAC 1196
 QY 1162 AATGCTCTCTTCAAAACCAAGATTCCTGAGAGTTCTCTAATGATGATGATGATG 1221
 Db 1197 CAGGCTCTCTTCAAAACCAAGATTCCTGAGAGTTCTCTAATGATGATGATGATG 1256
 QY 1222 GAGAGAGTGAACAGAACTTGTGAGAAATTAAGCTGAGATGATGATGATGATGATG 1281
 Db 1257 GAGCGGAGTGAACAGAACTTGTGAGAAATTAAGCTGAGATGATGATGATGATGATG 1316
 QY 1282 GAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1341
 Db 1317 GAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1376
 QY 1342 AAGGCGCTATCCAGAGACCGGCTAGTCTCCCTTACATGATGATGATGATGATGATG 1401
 Db 1377 AAGGCGCTATCCAGAGACCGGCTAGTCTCCCTTACATGATGATGATGATGATGATG 1436
 QY 1402 CACAGGAGTGAACAGAACTTGTGAGAAATTAAGCTGAGATGATGATGATGATGATG 1461
 Db 1437 CACGCGAGTGAACAGAACTTGTGAGAAATTAAGCTGAGATGATGATGATGATGATG 1496
 QY 1462 GAGTGAAGAGGCTCTTACAGAGATGAGATGATGATGATGATGATGATGATGATG 1521
 Db 1497 GAGTGAAGAGGCTCTTACAGAGATGAGATGATGATGATGATGATGATGATGATG 1556
 QY 1522 CCCCTCAGGCTGTGAGACCACTGATTTTACAGAGATGATGATGATGATGATGATG 1581
 Db 1557 CCCCTCAGGCTGTGAGACCACTGATTTTACAGAGATGATGATGATGATGATGATG 1616
 QY 1582 ACATCTGAGGCTGCGGATTTGATGATTTTCTGCTGCTCTTGGCCCCCAGCTTAT 1641
 Db 1617 GATCTGAGGCTGCGGATTTGATGATTTTCTGCTGCTCTTGGCCCCCAGCTTAT 1676
 QY 1642 TGTGTAATTTCTAATCTTGTGCTACTTGGTATGCTCAAAATTCATAATAGTCAACAG 1701
 Db 1677 TGTGTAATTTCTAATCTTGTGCTACTTGGTATGCTCAAAATTCATAATAGTCAACAG 1736
 QY 1702 TGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1761
 Db 1737 TGTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
 QY 1762 ACAAGAAATGCGCATCTTTTAAATATCTGAGTTTATTTACATTCAGAAATATGAGAG 1821
 Db 1797 ACAAGAAATGCGCATCTTTTAAATATCTGAGTTTATTTACATTCAGAAATATGAGAG 1856
 QY 1822 TGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1881
 Db 1857 TGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1916
 QY 1882 TTTCTGTGACAACTATCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1941
 Db 1917 CTTCTATGCTAATTCACCGCATGCTGCGCCATCACACCGAGGCTGCTGCTGCTGCTG 1976

QY 1942 CTGCCAGAGTGCATCTAGATTCACAAATGAACTTTCTGATTTGATTCATTTATTC 2001
 Db 1977 CTGCCAGAGTGCATCTAGATTCACAAATGAACTTTCTGATTTGATTCATTTATTC 2036
 QY 2002 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2061
 Db 2037 AGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2096
 QY 2062 GTCAACCATGCTGCGGAAATGAGAGT 2091
 Db 2097 GTTAAATGATGACAGGAGGAAAGGCTTATG 2126
 RESULT 9
 AF312713 2354 bp mRNA linear ROD 16-MAY-2001
 LOCUS AF312713
 DEFINITION Mus musculus sterolin (Abcg5) mRNA, complete cds.
 ACCESSION AF312713
 VERSION AF312713.2 GI:14091944
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2354)
 Lee, M.H., Lu, K., Hazard, S., Yu, H., Shulen, S., Hida, H.,
 Kojima, H., Allkimes, R., Sakuma, N., Pegoraro, R., Srivastava, A.K.,
 Salem, G., Dean, M., and Patel, S.B.
 Identification of a gene, Abcg5, important in the regulation of
 dietary cholesterol absorption
 Nat. Genet. 27 (1), 79-83 (2001)
 20578753
 MEDLINE 11138003
 PUBMED 2 (bases 1 to 2354)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submission
 JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 3 (bases 1 to 2354)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submission
 JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 Sequence update by submitter
 On May 16, 2001 this sequence version replaced gi:12382299.
 REMARK COMMENT
 FEATURES
 source
 1..2354
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /tissue_type="liver"
 1..2354
 /gene="Abcg5"
 139..2097
 /gene="Abcg5"
 /note="Abcg5"
 /note="ABCG5"
 /codon_start=1
 /product="sterolin"
 /protein_id="AA653097.1"
 /db_xref="GI:12382300"
 /translation="MEELPELSPEGRPHINRSLISLSDGSGVTGTEARHSLGLVLAH
 SYSVSRVGPWMNINSCQDKMDRLKDVSLYIESQIMCISLSSGSGRTTLLDAISG
 RLRRNTLEGEVFNCGELRDFODCFYVLOSIVFSLRVEIRLRTAMLAICRS
 SADVNRKVEAVTELSHVAADOMGSNFGSISGERRRVSIAOILQDPRKVMALD
 EPTGDKCMRTAMQIYLLAELNARRBIVYVTHORSELPQRPDKAIITLITELFCG
 TPBEMLGFNRNKGCPPEHSNPFDFMDLTSVDTGSREREIEFYRVMQLKCAFEED
 IYKILLENLERRARYLKTPTLPKTRKDPGMRGKGLVLRVTRNLRMRKQAVINHLV
 ONLIMGLFLIFILRYONNTLGAADVDRGLLYOLGVGATPYGMNANVLAIFNLRAVS
 DQSDGDIYHKQMMLAIVLAVLIPESVIATVIFSSVCWTGLLVEVNAFGRFSAALL

Mon Jul 28 09:43:27 2003

us-09-989-981a-5.rge

Page 13

BASE COUNT 573 a 604 c 594 g 583 t
ORIGIN
A P H I G C E E I L V V N E F G L I V O A N N I N V A S I A L L S I S L I G S G F I R I O E M P I P K L G F
T P K R C C E I L V V N E F G L I F T C G S N S M L N H P M C A I T O C Y Q I E K T C P G A T S E F T A N
F L I L G S I P A L V I L G I Y I F R V D I L I S R "

QY 2062 GTGAAGCCATGCTGGGAAATGAGAGTGC 2091
DB 2097 GTTAAGATGACGACGAGGAGGATTATG 2126

RESULT 10
AY195872
LOCUS
DEFINITION Mus musculus strain I/LmJ ATP-binding cassette sub-family G member 5 (Abcg5) mRNA, complete cds.
ACCESSION AY195872
VERSION AY195872.1 GI:31322255
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2351)
Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and Paigen, B.
TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/EI and I/Lm Inbred Mice
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2351)
Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
AUTHORS Direct Submission
TITLE Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA
JOURNAL
FEATURES
source
1.2351
/organism="Mus musculus"
/mol_type="mRNA"
/strain="I/LmJ"
/db_xref="taxon:10090"
/map="55 CM"
/sex="male"
/tissue.type="liver"
1.2351
/gene="Abcg5"
139..2097
/gene="Abcg5"
/note="ATP-dependent canalicular cholesterol transporter; white subfamily"
/product="ATP-binding cassette sub-family G member 5"
/protein_id="AA045093.1"
/db_xref="GI:31322255"
/translation="MGELPEFLSPEGARPHINRGSLSLEQSVTGTETARHSLGVLHV
SYSNRYGPMWNKSCOOKMDROILKDVSLITSEGLMCILGSSGSGKTYLLDLSIG
SADFTNKRVEAVTELSLASHADQVIGSNFGLSDVFLSVLRETLRYTAMALCRS
EPTTGIDMTANQIVLLAEALRDRIVYVTHOPRSELFQHFDAITLITGELVFCG
TPEMIGFNNKGYPCPEHNPEDFMDLSDVDSRREITRYRYOMLEAFESD
IYKILLENIERARYKTLPTVPEFKRDPGMRGKGLVLRTRTNLMNRKQAVIRLV
DNLMLGLFLFYLLRPNNTLKGAVQDVRGLIQFGVATPYTGMNANVLPMLRAVS
DOESODGLYHKKOMLAVYLAHPISIIATVIFSSVCWTGLYEPVAFGFSAAL
APHLIGELTLLVLSIVONPNIVSIVALLSGLISGFRINIOEMPIPLKIGYE
TFQRYCCELIVNVEYGLFTGCEBNTMLNHPKCAITQGVETIKETKPGATSRRTAN
FLLYGFPALVILGIVIFKVRDYLSR"

BASE COUNT 577 a 598 c 590 g 586 t
ORIGIN

Query Match 60.1%; Score 1406; DB 10; Length 2351;
Best Local Similarity 80.3%; Pred. No. 1.5e-295;
Matches 1662; Conservative 0; Mismatches 405; Indels 3; Gaps 1;

QY 25 CTGCACGGGCTCCCACTGAAGCACTCTGGGAGAGGTCGGCCACAGAAATTTGC 84
DB 57 CTCACATGGGCTCTCTCACTTAAGCTGCGCGAGCGACAGGCGACTAGAAATATCAC 116
QY 85 CCAAGCTTGTCTGCTGTGGCCATGAGTGAACCTCTCATCTTTGACCCCGGAGGTCAT 144

DB 117 TTGCATTTGCTTCCGCTAGACCATGGTGGAGTGCCTTTCTGAGTCCAGAGGACCCAG 176
QY 145 GGGTCTCCAACTAAACAGAGCTCCAGAGCTCCCTGGAGGGGCTCTCCACCGCCC 204
DB 177 AGGGCTCTACATCAACAGAGGGTCTCTAGCTCCCTGGAGGAGGTTGGTACAGGGGAC 236
QY 205 GGAGCCT---CAGAGCCGGGCATCTCCATCCCTCTCAAGCGCTCAAGCCAGCGGTAG 261
DB 237 GGAGGCTCGGACAGCTTAGTGTCTGCTGATGTCTCTTCAAGCGCTCAAGCAACCTGTGCG 296
QY 262 GCCCTGTGGACATCACTATCTTCCGAGCAGAGTGGACAGGAGATCTCCAAAGATGT 321
DB 297 GCTTGGTGGAACTCAATATCTCCAGCAAGATGGACAGGCAATCTCCAAAGATGT 356
QY 322 CTCTTTAGCTGAGAGCGGGGAGATCATGTGATCTTCAAGAGCTCAGGCTCGGGAA 381
DB 357 CTCTTGTATCATCAGAGGTGGCCAGATTATGTGATCTTAGGACAGCTCAGGCTCAGGGA 416
QY 382 AACACGCTGCTGAGAGCCATGTCCGGGAGGCTGGGGCGGGGACCTTCTCGGGGA 441
DB 417 GACACGCTCTGAGAGCCATCTCCGGAGGCTGGGACCTGGACCTGGMAAGGGA 476
QY 442 GGTGTATGTAGAGCGCGCGGCTGCGCGGAGACATTCAGAGACTTCTCTACGT 501
DB 477 CGTGTGTGAATAGGCTGGAGCTGCCAGAGGACCACTTCAGAGACTTCTCTACGT 536
QY 502 CTGCAAGAGCAACCTGCTGAGACAGCTTACCGGCGAGAGCGTCACTACACCG 561
DB 537 CTGCAAGAGCGAGCTTTTCTGAGACGCTCACTGTGCGGAGACGTTGGATACACAGC 596
QY 562 GCTGAGGCAATCGCGCGGCAATCCCGCTCTCTCCAGAAAGGTGAGGCGCTCAT 621
DB 597 GATGCTGGCCCTTGGCGAGCTCCCGGAGCTTCTCAACAAAGATGAGGACGATGT 656
QY 622 GGCAGAGCTGAGTCTGAGCATGTGCGACAGCCAGATGATGGCACTACAGCTTGGGGG 681
DB 657 GACAGAGCTGAGGCTGAGCAGCTGCGACAGCAAGATGATGGCACTATATTTTGGGG 716
QY 682 CATTTCCAGCGGTGAGCGCGCGGCTTCATCGGACCCAGCTGCTCAGGATCTTAA 741
DB 717 AATTCAGTGGCGAGCGCGCGGCTTCATCGGACCCAACTCTTCAGGACCCCA 776
QY 742 GGTATGCTGTTGATGAGCAACACAGGCTCGAGCTGATGATGATGATGATGAT 801
DB 777 GGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836
QY 802 CGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 861
DB 837 CTTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896
QY 862 CGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 921
DB 897 TCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 956
QY 922 TTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 981
DB 957 GTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1016
QY 982 TGAACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1041
DB 1017 TGAACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1076
QY 1042 GGAACGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1101
DB 1077 AGACCGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1136
QY 1102 AGCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1161
DB 1137 TGACATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
QY 1162 AATGCTCTTCAAAACCAAGATCTCTGAGTGTCTTCAATCGGCTGTCTCT 1221

Db	1197	CACGGTTCCTTTTAAACAAAAGATCCTCCTGGGATGTTCCGCAAGCTTGTCCTCTCT	1236
OY	1222	GAGGAGAGTGACACAGAAACTGCTGGAGAAATAGCTGGCAGTATTTACGCTCTCTTCA	1281
Db	1257	GAGGAGAGTAAACMAAACTTAATGAGAAATAGCAGGACAGTATTAATGCGTCCTCA	1316
OY	1282	GAAATGCATCAGGGTTGTTGCTCCTTTTCTGCTTCTGCGGGCTCCAGACATGCTCT	1341
Db	1317	GAATCTGATCAGGGCTCTCTCTCTCTATTTTCTACCTTCTCCGAGTCCAGAACACAGCT	1376
OY	1342	AAAGGCTCATCCAGGACCCGATGAGTCTCCTTAACAGATTTGTTGGGCGCACCCGTA	1401
Db	1377	AAAGGCTCTGTGCGAGAACCCGCTGGGCTGCTCATATGTTTGTGGGTGCACCCCTA	1436
OY	1402	CACAGGCACTCTGAACGCTGTGAATCTGTTCCGTGCGAGCTGCACGACAGCA	1461
Db	1437	CACGCGCATGCTCATGCTGTGAATCTGTTCCATGCTGAGAGCCGTCAGCGACAGGA	1496
OY	1462	GAGTCAGGAGCGCCCTACGCAAGATGGCBAATGATGCTGGCCTATGATCGACAGTCT	1521
Db	1497	GAGTCAGGATGGCTGTATCATATGATGGCAATGCTGCTCCCTACGCTACAGCGCCT	1556
OY	1522	CCCGTTCAGCCTGTGTGGCCACCATGATTTTACGAGTGTGTCATCTGAGAGCGGGCT	1581
Db	1557	CCCGTTCAGCATATCCGACAGGATTTTACAGAGTGTGTTATTTGATGCTGGGCTT	1616
OY	1582	ACATCTGAGGTTGCCGATTTGGATATTTTCTGCTGCTCTCTTGGCCCCCACTTAAT	1641
Db	1617	GTATCCCTGATGTTGCCAGATTTGGATATTTCTGCTGCTCTTTTGGCCCCCACTTAAT	1676
OY	1642	TGATGAATTTTAACTCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCAT	1701
Db	1677	TGGAGATTTTAAACACTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCAT	1736
OY	1702	TGTAGTGCTCTGCTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCAT	1761
Db	1737	TATAGTGCTCTGCTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCAT	1796
OY	1762	ACAAGAAATGCCCATCTCTTTTAAATCATCATGTTATTTTACATTCGCAAAATATTTGTCAG	1821
Db	1797	ACAAGAAATGCCCATCTCTTTTAAATCATCATGTTATTTTACATTCGCAAAATATTTGTCAG	1856
OY	1822	TGATATCTCTGTCATCATGATGTCATGAGCACTAATTTTACTTGTGGCAGCTCAATGT	1881
Db	1857	TGATATCTCTGTCATCATGATGTCATGAGCACTAATTTTACTTGTGGCAGCTCAATGT	1916
OY	1882	TTCTGTGTCATCATGATGTCATGAGCACTAATTTTACTTGTGGCAGCTCAATGT	1941
Db	1917	CACATGCTTAATACCCGATGTGCGCATATCCCAAGGGGTGAGTTTACGAGAAAC	1976
OY	1942	CTGCCCAGGTGCACATCTAGATTCACAAATGAACTTCTGATTTTGTATTCATTTATTC	2001
Db	1977	CTGCCCAGGTGCATCTAGATTCACAAATGAACTTCTGATTTTGTATTCATTTATTC	2036
OY	2002	AGCCTGTGATCTAGAGATAGTGTTCCTCAAAATTAAGGATCATCTCATTTAGCAGGTA	2061
Db	2037	AGCCTGTGATCTAGAGATAGTGTTCCTCAAAATTAAGGATCATCTCATTTAGCAGGTA	2096
OY	2062	GTGAAACCCATGCTGCGGAAATGAGTAGT	2091
Db	2097	GTTAAGATGACAGGCAAGAAAGGTTAATG	2126
RESULT 11			
AX320881			
LOCUS	AX320881	2258 bp	DNA
DEFINITION	Sequence 2 from Patent WO0179272.	linear	PAT 14-DEC-2001
ACCESSION	AX320881		
VERSION	AX320881.1	GI:17902431	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

[illegible]

OY 598 CCAGAGAGAGTGGAGGCGCTCATGCGAGAGCTAGTCTGAGCCATGTGGCAGACGACT 657
 DB 541 CAAAGAGAGAGTGGAGGCGCTCATGCGAGAGCTAGTCTGAGCCATGTGGCAGACGACT 600
 OY 658 GATTGGCACTACAGCTTTGGGGGCAATTTCCAGGGGAGAGGGGCGGGGCTTCATCCG 717
 DB 601 GATTGGCACTATATTTTGGGGGCAATTTCCAGGGGAGAGGGGCGGGGCTTCATCCG 660
 OY 718 AGCCAGAGCTCTCCAGAGATCTTAAGTCTAGTCTGTTTGAATGAGCCAAACAGAGCTGGA 777
 DB 661 AGCCAGAGCTCTCCAGAGATCTTAAGTCTAGTCTGTTTGAATGAGCCAAACAGAGCTGGA 720
 OY 778 CTGATGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 837
 DB 721 CTGATGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
 OY 838 TGTGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 897
 DB 781 TGTGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 OY 898 CATCTGAGCTGCTGAGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 DB 841 CATCTGAGCTGCTGAGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 OY 958 CAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
 DB 901 CAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 OY 1018 GAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1077
 DB 961 GAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 OY 1078 GATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
 DB 1021 GATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 OY 1138 AATGAAACACCTGAAACCTTACCAATGTTCTCTTCAAAACCAAGATTCCTGAGAT 1197
 DB 1081 AATGAAACACCTGAAACCTTACCAATGTTCTCTTCAAAACCAAGATTCCTGAGAT 1140
 OY 1198 TTTCTGAACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 DB 1141 GTTGGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 OY 1258 GGCAGTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
 DB 1201 GGCAGTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 OY 1318 TCTGCGGCTCCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
 DB 1261 TCTGCGGCTCCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 OY 1378 CCAAGTTTGGGGGCGCCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1437
 DB 1321 TCAAGTTTGGGGGCGCCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 OY 1438 GCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497
 DB 1381 GCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 OY 1498 GCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
 DB 1441 GCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 OY 1558 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
 DB 1501 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 OY 1618 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
 DB 1561 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620

OY 1678 CCAAAATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
 DB 1621 CCAAAATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 OY 1738 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1797
 DB 1681 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 OY 1798 TTTTACATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
 DB 1741 TTTTACATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 OY 1858 TTTTACATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
 DB 1801 TTTTACATTCAAATATATGATCAACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 OY 1918 AGGATATTCATTCATGAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
 DB 1861 AGGATATTCATTCATGAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 OY 1978 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2037
 DB 1921 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 OY 2038 AAGGATCATCTGATTCATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 2091
 DB 1981 AAGGATCATCTGATTCATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 2034
 RESULT 12
 AF312714 2470 bp mRNA linear ROD 26-AUG-2002
 LOCUS Rattus norvegicus sterol (Abcg5) mRNA, complete cds.
 DEFINITION AF312714
 ACCESSION AF312714.3 GI:22477143
 VERSION
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 2470)
 Lee, M.H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hieda, H.,
 Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R., Silvestra, A.K.,
 Salen, G., Dean, M. and Patel, S.B.
 Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption
 Nat. Genet. 27 (1), 79-83 (2001)
 JOURNAL MEDLINE
 PUBMED 20578753
 REFERENCE 11138003
 2 (bases 1 to 2470)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submision
 Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 3 (bases 1 to 2470)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submision
 Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 4 (bases 1 to 2470)
 Lu, K., Lee, M. and Patel, S.B.
 Direct Submision
 Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 REMARK
 COMMENT Sequence update by submitter
 On Aug 26, 2002 this sequence version replaced gi:14091945.
 FEATURES
 location/Qualifiers
 1..2470
 source

/gene="Abcg5"
 /note="ABCG5"
 /codon_start=1
 /product="sterolin"
 /protein_id="AAG3098.3"
 /db_xref="GI:22477144"
 /translation="MSLPEFLSPGARGPHNNRGSQSSLEBGSVYSGSEARHSRLGLNV
 SFSSNRVAGPMWNKSCQOKDKRLILDVSLYIESQGTMLTSSGSGKTTILDAISG
 RLRTGTGLGEVFNCGCELDRDQFODCVSLQGDVLSLTIVETRLRYAMLAIRSS
 SADYDKRVEALTELSLHVAQMDICNNGSGISGERRVSAALQLODPKYMID
 EPTGLDCCMNIHIVLLLELARNRRIYIYTIHQPSELHPHPDKITLYEGELVFCG
 TPTEMLGFENNCGYCPDEHSNPDEFDVLDVDSQNSRETIYTKRYQMLAESRQSD
 ICHILINIEIRHLIKTLPMVPEKTKPDMFCFCLGVLLETRVTLNRNNOVIMRLY
 ONILMGFLIETYLIRVONNMHLKGAQDRVGLTQLVGATVYRLMLNRAVNPMLRAVS
 DQESQDGLYOKQMLNLAIVLHALPESTVAIVTSVYGMVLGIEPVARGYFSALL
 APHLHIGELTILVIGWQNPINIVSLTISGLTGSQFINIEMEPDLKILGFE
 TFQKQCELTIVNEFYGLNFTGGSNYSVNNPMGCSMTQGIQPLEKTCPCATSRFTTN
 FLIYLSYCEPILVILGMYEVRDYLLSR"

Query Match	Best Local Similarity	59.1%	Score 1383.8	DB 10	Length 2470
Matches 1641	Conservative	0	Mismatches 407	Indels 3	Gaps 1
QY	44	TGAAGCCACTCTGGGGAGGGGTCCGGCCACCAGAAAATTTGCCAGCTTTCCCTCGTTG	103		
Db	2	TAAAGTTCCTGTGAAGCCAGACAGAACCCAGAGANTCTACATTTAGTCCCGCTG	61		
QY	104	GCCATGGGTACACCTCTTCATCTTTGACCCCCCGAGGGTCCATGGGTCTCCAAGTAAACAG	163		
Db	62	GCCATGTAGTAGCTGCCCTTTCTTGAGATCCAGAGGGAGCCAGAGGGGCTCCACAAACACAG	121		
QY	164	GCGTCCAGAGCTCCCTGGAGGGGGTCCCGCCACCGCCCGGAGGCT---CACAGCCTG	220		
Db	122	GGGTTCAGAGCTCCCTGGAGGAAGGCTCAGTTACAGGCTCAGAGGCTCGGCACACCTTA	181		
QY	221	GGCATCTCCATGCTCTTACAGCGTCAGCCAGCCAGCGGTGAGGGCTCTGTGGACATCACA	280		
Db	182	GGTGTCTGATGTGTCTTACAGCTCAGCAACCGTGTGGGGCTTGAGGAACATCAA	241		
QY	281	TCGTGCGGGAGAGTAGGACAGGCGAGTCCTAAAGATGTCTCCCTTTAGCTGGAGAC	340		
Db	242	TCATGCCAGAGAAAGTGGACAGAAATTCCTAAAGATGTCTCTTGTACATCGAGACT	301		
QY	341	GGCGAGATCATGTGATCTCTAGGAACCTAGGCTCCGGGAAAAACAGCTGCTGGAGCGC	400		
Db	302	GGCGAAGCATGTGCATTTAGTACTAGGCTAGGGAAAAACAGCTGCTGGAGCGC	361		
QY	401	ATGTCCGGAGAGCTGGGGGGCGGGGACCTTCTGGGGAGGTATGTGAAGGGCGG	460		
Db	362	ATCTCTGGGAGGCTGGGGGACAGAGGACCTTGAAGGAGGTGTTGTGAAGGGCTGC	421		
QY	461	GGCGTCCGCGGAGAGCATTTCCAGAGACCTCTCTCTTACGTCTCGAGAGCCAGACCCG	520		
Db	422	GAGCTCGGAGGAGACAGTTCCAGACAGCTGCTCTTACCTCTCGAGAGCCATGTCTTT	481		
QY	521	CTGAGAGGCTTACAGCTGCGGAGAGAGCGTGCACACCGCGCTCTGGCCATCCGCGC	580		
Db	482	CTGAGAGGCTTACAGCTGCGGAGAGAGCGTGCATACAGCGGAGTCTGGCTTCGGAGC	541		
QY	581	GGCAATCTCGGGGCTCTCCAGAGAAAGTGGAGGGCGGCTATGGCAGAGCTGATGTGAC	640		
Db	542	AGCTCCGGAGACTTTACAGCAAGAAAGTAGAGGAGTCTCTCAGAGACTGATGTGAGC	601		

QY	641	CATGTGGACACCGACGATGGATTGGCACTACAGCTTGGGGGCAATTTCCACGGGTGAGCGG	700
QY	641		
Db	602	CACGTGGCAACCAAAATGATGGCAACTTAATATTTGGGGGGAATTTTCCAGTGGGACGGG	661
QY	701	CGCGGGGTCCATCCGACAGCCCGAGTGCATCCAGGATCCCTAAGGTGATCGTGTATTATGAG	760
QY	701		
Db	662	CGCGAGATGCCATCCGACGCCCACTCTTTCAGAGACCCCAAGGTATGATGGTTGATACGAG	721
QY	761	CCAAACCAAGCGCTGGACTGCATGACTGCTAATTCAGATTGTGTCTCTCTGTTGGATGACG	820
QY	722	CCAAACCAAGAGATGGAGCTGCATGACTGCAAAATCATATGTCTCTCTCTGTGGTCAAGCTG	781
QY	821	GCATCGAGAGACCGGATTTGGTGTCTACCACTTACACACCACCCCGTTCTGAGCTTTTTCAG	880
QY	782		
Db	881	CTCTTTGACAAAATTTGCCATCCGACCTCGGAGAGCGATTTCTGTGGGACGGCACGG	940
QY	842	CACCTTGGACAAAATTTGCCATTTCTGCACTTACGAGAGATGGTGTGTGGGACCGCACAG	901
QY	941	GAATTCCTTGATTTCTTCAATGACTGCGGGTTACCCCTTGCTCGAATTTCAACCCCTTTT	1000
QY	902	GAGATCTCGGGCTTCTTCAATTAAGTGGTATACCCCTGCTCGAATTTCCAAATCCCTTT	961
QY	1001	GATTTCTAATGTGACTGACGCTCACTGGATGGATACCCCAAGACAGGAGGAAATAGAAACC	1060
QY	962	GATTTCTAATGTGACTGACATCGGTGGACACCCCAAGACAGGACGAGATAGAGACG	1020
QY	1061	TCCAAAGAGAGTCCAGATGATAGAAATCGCTACAAAGAAATGCAATTTGTTCATTAATCT	1120
QY	1022	TACAAACGAGTCCAAATGTCTGGAATCTGCTTCAAGGCAATGGACATCTGTCACAAAATC	1081
QY	1121	TTGAGAAATATTTGAAAGAAATGAAACACCTGAAAGATTAACCAATGGTTCCTTTCAAAAC	1180
QY	1082	CTGGAGAACATTTGAAAGAAAGAACAGACCTGAAACCCTAACCATGTCTTCAAAACG	1141
QY	1181	AAAGATTCCTCCGAGATTTTCTCTAATCGGTGTCTCTGTAGAGAGTGCACAAAGAC	1240
QY	1142	AAAAATCTCCCGGAGATTTCTGTGACAGCTCGGGCTTCTCCGAGAGAGTATACGAGAAC	1201
QY	1241	TTGTGTAGAAATTAACCTGCAAGTGTATACGGTGTCCCTGCAATTCGATTCATGAGGTTTG	1300
QY	1202	CTAATGAGGAATTAACCAAGTGTGTATTAATGAGTCTTGTGTGCAATTCGATTCATGAGGCTG	1261
QY	1301	TTCTCTCTTTCTTCTGTTCTCGCGGGTCCGAAGATGCTTAAAGGGTGTATCCAGAC	1360
QY	1262	TTCTCAATTTTCTACCTTCTCGAGTCCAGAAACAACATGCTGAAGGGCGCTGTTCAGACG	1321
QY	1361	CGCGTAGTCTTCCTTACCAATTTGTGTGGCGCCACCCCGTACACAGGACGATCTGAACGCT	1420
QY	1322	CGCGTAGGAGCTGTGTGTACCACTTGTGGGTGCCACCCGTTACACCGGACGATCTGAACGCT	1381
QY	1421	GTGAATTCGTTTCCGTCGCGAGCGTGCAGCGACACAGAGATCAAGACGGCTCTTAC	1480
QY	1382	GTGAATTCGTTTCCGTCGAGAGCTGTGCACGACACAGAGAGATCAAGATGTGCTGTAC	1441
QY	1481	CAGAAGTGGCAGATGATGCTGGCTATGACATGAGCAGCTCCGCCCTTACACGCTTGTGTCC	1540
QY	1442	CAGAAGTGGCAGATGATGCTGGCTATGACATGAGCAGCTCCGCCCTTACACATGCTGTCC	1501
QY	1541	ACCATGATTTTCAAGAGTGTGTCTACTGACAGCTGGGCTTACATCTCGTAGTGTGCCGA	1600
QY	1502	ACGGATGATTTTCAAGAGTGTGTACTGTGACTGTGGCTTGTATCCGAGGTGCCACA	1561
QY	1601	TTTGGATTTTCTTGTGCTGTCTTGTGGCCCCCACCCTTAATTTGGTGAATTTCTAATCTT	1660
QY	1562	TTTGGATTTTCTTGTGCTGTCTTGTGGCCCCCACCCTTAATTTGGAGAAATTTTGCACACTT	1621
QY	1661	GTGTACTTGTGTATGTCACAAAATCCAAATATGTCAAAGTGTACTGGCTGTGCTGTCC	1720
QY	1622	GTGTGTGTGTATATGTCACAAAATCCCAATATTTGTCAACGATATATGTGTGTGCTGTGAGT	1681
QY	1721	ATTCCGGGGGCTCTTGTGATCTGATTCCTTCAGAAACATACAAAGAAATGCCCATTTCT	1780

1184 GATTCTCGAGAGTTTCTCTAACTGGGCTTCTCTAGAGAGCTGACAGAACTTG 1243
1081 GATCCCTCGGGAGTGTGGCAAGCTTGGTCTCTAGAGAGCTGACAGAACTTG 1140
1244 GTGAGAAATAGCTGAGTGAATAGCGCTCTCTCTAGAGAGCTGACAGAACTTG 1303
1141 ATGAGAAATAGCTGAGTGAATAGCGCTCTCTCTAGAGAGCTGACAGAACTTG 1200
1304 CTCCT 1363
1201 CTCATCT 1260
1364 GTAGAGCT 1423
1261 GTGGGGCT 1320
1424 AATCT 1483
1321 AATCT 1380
1484 AAGTGGCAGAGTCT 1543
1381 AAGTGGCAGAGTCT 1440
1544 AATCT 1603
1441 GTCATCT 1500
1604 GATCT 1663
1501 GATCT 1560
1664 CT 1723
1561 CT 1620
1724 GCGGGGCT 1783
1621 CT 1680
1784 AATCT 1843
1681 AATCT 1740
1844 TTTACGAGCTGAAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1903
1741 TTTACGAGCTGAAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800
1904 TGTGCT 1963
1801 TGTGCT 1860
1964 TTTCAAAATGAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2023
1861 TTTCAAAATGAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920
2024 GTTGTCT 2087
1921 GTTGTCT

RESULT 14
AX456526
LOCUS AX456526
DEFINITION Sequence 48 from Patent WO0227016.
ACCESSION AX456526
VERSION AX456526.1 GI:21715414
KEYWORDS
ORGANISM synthetic construct
SOURCE synthetic construct
REFERENCE 1
AUTHORS Patel,S.B. and Dean,M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
JOURNAL Patent: WO 0227016-A 48 04-APR-2002;
THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel, Shalendra B. (US) ; Dean, Michael (US)
FEATURES
source 1. 2035
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT 481 a 533 c 537 g 484 t
ORIGIN
Query Match 58.2%; Score 1363; DB 6; Length 2035;
Best Local Similarity 80.6%; Pred. No. 3.5e-286;
Matches 1607; Conservative 0; Mismatches 385; Indels 3; Gaps 1;
100 GTTGGCATTGGGTGACCTCTCATCTTTGACCCCCGGAGGTCATGGGTCTCCAGTAA 159
1 GCTGGCATTGGGTGACCTCTCATCTTTGACCCCCGGAGGTCATGGGTCTCCAGTAA 60
160 CAGAGGCTCCAGAGCTCCCTGAGAGGGGCTCTCCAGAGCCCGGAGGCT---CACAG 216
61 CAGAGGCTCCAGAGCTCCCTGAGAGGGGCTCTCCAGAGCCCGGAGGCT---CACAG 120
217 CCTGGGCT 276
121 CTGAGGCT 180
277 CACATCT 336
181 CAATCT 240
337 GAGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 396
241 GAGTGGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
397 CGCCATGTCGGGAGGCTGGGGCGGGGAGGCTCTCTCTCTCTCTCTCTCTCTCT 456
301 CGCCATGTCGGGAGGCTGGGGCGGGGAGGCTCTCTCTCTCTCTCTCTCTCTCT 360
457 CGGGGCGGCTGGGGCGGGGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 516
361 CTGGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
517 CTTGCTGAGAGCT 576
421 CTTGCTGAGAGCT 480
577 CCGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 636
481 CAGCAGCTCCCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
637 GAGCAGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 696
541 GAGCAGCTCCCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
697 GCGGCGGCGGCT 756
601 GCGGCGGCGGAGTCT 660
757 TGAGCCAGCAGAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 816
661 CGAGCCAGCAGCAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 720
817 ACTGGCT 876
721 GCTGGCT 780
877 TCAGCT 936
781 CCAGCAGCT 840

Qy	937	AGCGAAATCTTGATTTCTTCATGATCAGCGGGTTACCTTGTCCGAAATTCAAACC	996
Db	841	AGAGAGAGATGCTCGGCTCTTCAATPACGTGGTTAACCCCTCTCTGAACATTCAAATCC	900
Qy	997	TTTTGACTCTAATGAGACCTGACGTCACTGATACCCAAAGCAAGAAACGGAAATAGA	1056
Db	901	CTTTGATTTCTACATGAGACTTGACATCGGTGGACACCAAACAGAGAGCGGAGATAGA	960
Qy	1057	AACTCCAAAGAGAGTCCATGATGAAATCTCCCTCACAAGAAATAGCAATTTGTCATAA	1116
Db	961	GAGCTACAAAGCGATGCCATGATCTGGAAATCTCCCTTCAGGCAATGGACATCTGTCAAA	1020
Qy	1117	AACTTGAAGAAATATGAAAGAAATGAAACACCTGAAAAGCTTAACCAATGGTTCCTTCA	1176
Db	1021	AAATCCGAGAAACATTTGAAGAAACAAGACACTGAAAACCTTACCATGGTTCCTTCAA	1080
Qy	1177	AAACAAAGATTTCTCTCGAGTTTCTTCTPAACTGGGTCTCTCTGAGAGAGTGAACAAG	1236
Db	1081	AAAGAAAATTCCTCCCGAAATGTTCTGCAAGCTCGGGCTTCCCTGAGGAAATGACAG	1140
Qy	1237	AAACTGGGAGAAATTAAGCTGGGACGTATTAACGGTCTCCTTGAATCTGATCATGG	1296
Db	1141	AAACCTAATGAGAAATTAAGCAGGTGGTATTAATGCTTTCTTGAATCTGATCATGG	1200
Qy	1297	TTTGTTCCTCTTTCTCTGTTCTGTCGGGGTCCGAAGCAATGGCTAAAGGGTGTATCCA	1356
Db	1201	TCGTGTCCTCATTTTCTTCACTTCTCCGAGTCCAGAACACATGCTGAAGGGCTGTCCA	1260
Qy	1357	GGACCGCTAGTCTCTTTTACAGATTGTGGGGCCACCCCTGACACAGCATGCTGAA	1416
Db	1261	GGACCGCTAGGGCTGTGTACACAGCTTGTGGGTGCCACCCCGTACACCGCATGCTCAA	1320
Qy	1417	CGCTTTGGAATCGTTTCCCGTGCAGAGCTGTACAGCACAGAGATCAGAGCGGCT	1476
Db	1321	CGCTGTGAACCTCTTCCCATGTGAGAGCTGTACGACAGAGAGATCAGAGATGGCT	1380
Qy	1477	CTACCAAGATGGAGATGATGTGCGGCTATGATGACATGACGTCTCCCTTACAGCGTTG	1536
Db	1381	GTACCAAGATGGAGATGATGTGCTCGGCTATGATGATGATGATGATGATGATGATGATG	1440
Qy	1537	TGCCAATGATTTTACAGAGTGTGTCTACTGAGACGCTGGGCTTACATCTAGGTTGC	1596
Db	1441	TGCCAAGTGTATTTACAGACGCTGTACTGTACTGTACTGTGGGCTGTATGCCAGGTGCC	1500
Qy	1597	CCGATTTGGAAATTTTCTGCTGCTCTCTTGGGCCCCCACTTAATTTGGATTTCTAAAC	1656
Db	1501	CAGATTTGGAAATTTTCTGCTGCTCTCTTGGGCCCCCACTTAATTTGGAAATTTCTGAC	1560
Qy	1657	TCATGCTACTTGTGTATGCTTCCAAATTCAAATATAGTAAAGTATGGTCTGTGCT	1716
Db	1561	ACTTGTCTGCTGTGTATGCTTCCAAATTCAAATATAGTAAAGTATGGTCTGTGCT	1620
Qy	1717	GTCATTTGGGGGGGCTTGTGTGATGTGGATTCAGAAACATCAAGAAATGCCAT	1776
Db	1621	GAGATTTTGGGGGGGCTTGTGTGATGTGGATTTATCAGAAACATCAAGAAATGCCAT	1680
Qy	1777	TCCTTTAAATATCATGATTTTACATTTCCAAAATATATGACGTGAGATTTCTGTACT	1836
Db	1681	TCCTTTAAATATCTGGGTTACTTACCTTCCAAAATATATGATTTCTGTGTACT	1740
Qy	1837	CAATGACTTCTACGAGCTAATTTTCACTTGTGACACTCAAAATGTTCTGTGACAACATA	1896
Db	1741	CAATGAGTTCTATGAGCTGAACTTCACTTGTGTCCTCAACACTTCTGTGCCAAATATA	1800
Qy	1897	TCCAATGTGCTCTCACTCAGGAATTCATTTCAATTTGAGAAAACCTGGCCAGGTGCAC	1956
Db	1801	CCCAATGTGCTCTCACTCAGGAATTCATTTCAATTTGAGAAAACCTGGCCAGGTGCAC	1860
Qy	1957	ATCTAGATTCACATATGATTTTGTATTTGTATCATTTATTTCCAGCTCTTGTATCTCT	2016
Db	1861	GTCAGATTTACACAAACTCTCTGATCTTGTACTGTGTACATCCGCACTCTTGTATCTCT	1920

RESULT 15	AX456523	1915 bp	DNA	Linear	PAT 06-JUL-2002
LOCUS	AX456523	Sequence 45 from Patent WO0227016.			
DEFINITION	AX456523	GI:21715412			
ACCESSION	AX456523				
VERSION	AX456523.1				
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
BASE COUNT	453 a	502 c	484 g	476 t	
ORIGIN					
Query Match	57.1%;	Score 1335.8;	DB 6;	Length 1915;	
Best Local Similarity	81.5%;	Pred. No. 2.9e-280;			
Matches 1560;	Conservative	0;	Mismatches 352;	Indels 3;	Gaps 1;
107	ATGGGTGACCTCTCTATCTTTGACCCCGGGAGGGTCCATGGGCTCCAGTAAGACAGAGCC	166			
1	ATGGGTGACCTCTCTATCTTTGACCTCCAGAGGAGCCAGAGGCTCACAATCACAGAGGG	60			
167	TCCAGAGCTCCCTGAGAGGGGGCTCCAGCCAGCCCGCCGAGGCT---CACAGCCCTGGCC	223			
61	TCTTGAGCTCCCTGAGAGAGGTTCGGGTACAGGGCACAGAGCTCGGACACGCTTAAGT	120			
224	ATCTTCATGCTCTCTTACAGCGCTCAGCCACCGGTGAGGCTCTGGTGGACATCATCT	283			
121	GTCTCGATGCTCTCTACAGCGCTCAGCAACCGTGGGCTTGGTGAACATCAATCA	180			
284	TGCGGCGAGCTGGAGACCGAGATCTCAAGATGTCCTCTTGACTGGAGAGCGGG	343			
181	TGCCAGCAGAGTGGACAGCAATCTTAAAGATGTCTCTTGAATCAATCAGAGTGGC	240			
344	CAGATCATGTGATCTAGGAAGCTCAGGCTCCGGGAAAAACACAGCTGCTGGACGCATG	403			
241	CAGATTTATGTCACTTAAAGCAGTCAAGGTGAGGAAACACAGCTGCTGGACGCATC	300			
404	TCCGGAGGCTGGGGGCGCGGGACCTTCCTGGGGAGGTATGTGAACCGCCGGCG	463			
301	TCCGGAGGCTGGGGGCGCGGGACCTTCCTGGGGAGGTATGTGAACCGCCGGCG	360			
464	CTGGCGCGGAGAGCTTCCAGAGCTCTCTCTTACGCTCTGACAGAGGACACCTGCTG	523			
361	CTGGCGGAGGACCACTTCCAGAGCTCTCTCTTACGCTCTGACAGAGGACCTTCTGCT	420			
524	AGCAGCTCAACCGTGGCGGAGAGCGTCACTACACCGCGCTCTGGCCATCCGCGCGGC	583			
421	AGCAGCTCACTGTGGCGGAGAGCGTTCGATACACGAGATGCTGGCCCTCTGCGCGAC	480			
584	AATCCGGGCTCTTCCAGAAAGATGGAGGCGCTCATGCGAGACTGATGAGCCAT	643			


```
Db 481 TCCGGGACTTCTACAAAGAGGTAGAGGACATGACAGAGCTGAGCCCTGAGCCAG 540
Qy 644 GTGGGAGACCGAGATGATTGGCAACTACAGCTGGGGGCAATTTCCAGGGTGAAGGGGCG 703
Db 541 GTGGGGAGACCAAAATGATTGGGAGCTATATTTTGGGGAAATTTCCAGTGGGAGGCGGCG 600
Qy 704 CGGGCTCCATGCGAGCCAGCTGCTCCAGATCTTAAGTCAAGCTCTTTGATGAGCCA 763
Db 601 CGAGTTCCATGCGAGCCCAACTCTCTCAGGACCCCAAGATCATGATGATGAGCA 660
Qy 764 ACCAGGCTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 823
Db 661 ACCAGGAGCTGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
Qy 824 CGCAGGAGCCGAATTTGTTCTCAGATTCACAGCCGCTTCTGAGCTTTTTCAGCTC 883
Db 721 CGCAGGAGCCGAATTTGTTCTCAGATTCACAGCCGCTTCTGAGCTTTTTCAGCTC 780
Qy 884 TTTGACAAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 943
Db 781 TTTGACAAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Qy 944 ATGCTGATTTCTTCAATGAGTGGGCTTACCTTGTCTGATGATGATGATGATGATGAT 1003
Db 841 ATGCTGATTTCTTCAATGAGTGGGCTTACCTTGTCTGATGATGATGATGATGATGAT 900
Qy 1004 TTTTATATGAGCTGAGCTGAGTGAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1063
Db 901 TTTTATATGAGCTGAGCTGAGTGAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy 1064 AAGAGATGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1123
Db 961 AAGAGATGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
Qy 1124 AAGAAATATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 1183
Db 1021 GAGAAATATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 1080
Qy 1184 GATTCCTCTGAGTGTCTCTAAACCTGGGCTTCTCTGAGGAGAGAGAGAGAGAGAG 1243
Db 1081 GATTCCTCTGAGTGTCTCTAAACCTGGGCTTCTCTGAGGAGAGAGAGAGAGAGAG 1140
Qy 1244 GTGAGAAATGAGCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1303
Db 1141 ATGAGAAATGAGCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1200
Qy 1304 CTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1363
Db 1201 CTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Qy 1364 GTAGGCTCTCTTACAGAGTTTGGGGGCGACCCCGTACAGAGAGAGAGAGAGAGAG 1423
Db 1261 GTAGGCTCTCTTACAGAGTTTGGGGGCGACCCCGTACAGAGAGAGAGAGAGAGAG 1320
Qy 1424 AATCTGTTCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1483
Db 1321 AATCTGTTCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Qy 1484 AATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1543
Db 1381 AATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
Qy 1544 ATGATTTTACAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1603
Db 1441 GTGATTTTACAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Qy 1604 GATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1663
Db 1501 GATATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
Qy 1664 CTGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1723
Db 1561 CTGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
```

Search completed: July 27, 2003, 05:52:29
Job time : 8322.43 secs

```
Qy 1724 GCGGGGGTGTGTTGATCTGATCTCTGAGAAACATACAGAAATGCCATTCCTTTT 1783
Db 1621 TCTGGGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Qy 1784 AAAATCATGATTTTATTTTACATTCGAAATATTTGAGTGGATTTCTGTGATGATGAG 1843
Db 1681 AAAATCCGGGTTATTTTACATTTCCAAAATATTTGATGATGATGATGATGATGATG 1740
Qy 1844 TTTACGAGCTGAAATTTTCACTTGTGAGAGCTCAATGTTCTGTGATGATGATGATG 1903
Db 1741 TTTACGAGCTGAAATTTTCACTTGTGAGAGCTCAATGTTCTGTGATGATGATGATG 1800
Qy 1904 TGTGCTTCACTCAAGGATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 1963
Db 1801 TGTGCTTCACTCAAGGATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTT 1860
Qy 1964 TTTCAATGAGACTTTCTGATTTTGTATTTATTTTCAAGCTCTGTGATGATGATG 2018
Db 1861 TTTCAAGGAGAACTTCTCATCTTATATGAGTTTATCCAGCTGTGATGATGATG 1915
```


US-09-252-991A-13705
; Sequence 13705, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13705
; LENGTH: 1668
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13705

Query Match 2.4%; Score 57; DB 4; Length 1668;
Best Local Similarity 45.7%; Pred. No. 1.1e-05;
Matches 238; Conservative 0; Mismatches 280; Indels 3; Gaps 1;

QY 312 TCMAAGATGCTCTCTGACGTGAGAGCGGCGAGATCATGTCATCTAGAGACTGAG 371
DB 959 TCGACGGGCTCAATTTCAGTACACCGCGGCGAGACCTGGGCACTTGGCGAAACGG 1018
QY 372 GCTCGGGAACCAACCGCTGCTGACGCGCATGTCGCGAGAGCTGGGGCGCGGAGCT 431
DB 1019 GCTCGGGAACCAACCGCTGCTGACGCGCATGTCGCGAGAGCTGGGGCGCGGAGCT 431
QY 432 TCCTGGGGAGGTGATGTAAGCGCGCGGCGCTGCGCGGAGAGAGTTCAGAGCTGCT 491
DB 1079 TCCGCTTCGAAAGGACCGCGCTGAGCGGTCTGCGCGCAACATGACGTGCGCGCTGCGC 1138
QY 492 TCTCTACGCTCTGACAGACGACACCGTGTGAGCAGCGCTGACCGGTGCGCGAGAGCTGCG 551
DB 1139 GCGAGATGACAGGTGTGTCAGAGACCATATGACAGCTTACGCCACGCAATGTGTGCG 1198
QY 552 ACTACACCGCGCTCTGCTGACATCCCGCGGCAATCCCGGCTCTCTCTGAGAGAGTGG 611
DB 1199 GCGAGATGCTGCGGAGAGCGCTGCGCATTCATGAGATGCGGAGAGAGCGGCGCAAGAGAG 1258
QY 612 AGCGCGTATGAGCAGAGCTGATGAGCCATGTGCGAGAGCGGCTGATGTGGCACTGACA 671
DB 1259 AGCGGATCATGACGCGCTG---GTGAGGTGCGGGCTGATTCGCGACAGACCGGCTACCGTT 1315
QY 672 GCTTGGGGGCAATTCACGCGGTGAGCGCGGCTGCTGATGCGAGCGGCGGAGTGTGTC 731
DB 1316 ACCCGGCAATTCGCGGGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1375
QY 732 AGGATCTAAGGATGCTCTTTGATGAGCGCAACCGAGCGCTGAGCTGATGCTGCTGCTA 791
DB 1376 TGAACCGGCGGATGCTGCTGCTGAGCAACCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGG 1435
QY 792 ATCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 832
DB 1436 GCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476

RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.4%; Score 56.8; DB 3; Length 4403765;
Best Local Similarity 46.0%; Pred. No. 0.0009;
Matches 273; Conservative 0; Mismatches 312; Indels 9; Gaps 2;

QY 271 GGACATCATCTTCCCGGACAGTGTGACGACGAGATCTCAAGATGCTCTGTA 330
DB 2879474 GGTCTGAGATTCAGAGGCGGGGTACCGCTGCGCGGATGACGCGTTAAGCTGCA 287953
QY 331 CGTGAGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 390
DB 2879534 CGTGAGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 390
QY 391 GCTGAGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
DB 2879534 CGTGAGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
QY 451 GAACGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 510
DB 2879654 CGACATCAGACGCGTGAAGGCGCGCGCTGCGGAGATGAGCTGAGATGAGAT 287971
QY 511 CGACATCAGACGCGTGAAGGCGCGCGCTGCGGAGATGAGCTGAGATGAGATGAGATGAGAT 570
DB 2879714 CGTCTTCAGAGCGTGAAGGCGCGCGCTGCGGAGATGAGCTGAGATGAGATGAGATGAGAT 570
QY 571 CATCGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 630
DB 2879774 GCTGCGGCGGCGG---CGTGTACGAGCGCGCGCGGTGAAGGCTGCGGAGAGCTGCT 2879831
QY 631 GAGTGTAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 690
DB 2879831 GATCGGAGTCAATTCGCGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 690
QY 691 GGTGAGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750
DB 2879885 CGGCGGAGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750
QY 751 GTTGATGAGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 810
DB 2879945 TCGCGGAGCGGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 810
QY 811 GGTGAGCGGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
DB 2880005 CCGCGGAGCGGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864

RESULT 6
US-09-103-840A-1
; Sequence 1, Application US/09103840A

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A

us-09-989-981a-5.rni

US-09-252-991A-9760

Query Match	2.48;	Score 56.2;	DB 4;	Length 1713;
Best Local Similarity	45.68;	Pred. No. 1.8e-05;		
Matches 276;	Conservative	0;	Mismatches 323;	Indels 6;
			Gaps 2;	

[illegible]

RESULT 9
 US-09-252-991A-10208/C
 : Sequence 10208, Application US/09252991A
 : Patent No. 6551795
 : GENERAL INFORMATION:
 : APPLICANT: Marc J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 : FILE REFERENCE: 107196, 136
 : CURRENT APPLICATION NUMBER: US/09/252,991A
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US 60/074,788
 : PRIOR FILING DATE: 1998-02-18
 : PRIOR APPLICATION NUMBER: US 60/094,190
 : PRIOR FILING DATE: 1998-07-27
 : NUMBER OF SEQ ID NOS: 33142
 : SEQ ID NO 10208
 : LENGTH: 2805
 : TYPE: DNA
 : ORGANISM: Pseudomonas aeruginosa
 : US-09-252-991A-10208

Query Match	2.48;	Score 56.2;	DB 4;	Length 2805;
Best Local Similarity	45.68;	Pred. No. 2.4e-05;		
Matches 276;	Conservative	0;	Mismatches 323;	Indels 6;
				Gaps 2;

QY 311 CTCGAAGATGTCCTCTTGTACGTGAGACGGGACAGATCATGTGATCTCTAGGAAGCTCA 370
 Db 1629 CTCGAAGGGGGTATCGTTCCAACTGTGAAGCCGGCAGAACGCTCGCTGGTGGCGATCC 1570
 QY 371 GGCTCCGGCAAAACACAGCTGCTGAGCCGATGTCGCGGAGGCTGGGGGGGGGGAGAC 430
 Db 1569 GGGTGGGGCAAGTCGACCTGGCGGGCGCCCTTGACCTGATGAGGAACCACTCCGGC 1510
 QY 431 TTCTCTGGGGAGGTGATGTGAACGGCCGGGCGCTGGCCCGGAGCACTTCAGAGTCG 490
 Db 1509 TGGCTGAAAAATGCGCGGGGAGGAGTCAAGGGCGGCACAGAGACACGCGCGGCA---G 1453
 QY 491 TTCTCTACGTCCTCCAGAGCAGACACCTTGCTGAGCAGACGTCACACGTCGGGAGACGTCG 550
 Db 1452 TTGCGCGCGGAGCTGCAAGATGTCCTTCCAGAAACCCCTACGCTCGCTCAATCCGCGACG 1393
 QY 551 CACTACACCGCGCTCTGGCCATCCGCGCGGCAATCCCGGCTCTCTTCAGAAAGAGTG 610
 Db 1392 AAGATCGCGGACCACTGTGGCGAGCGCTGCTGATCAACACCGGCGTGTGCGGGAGGAA 1333
 QY 611 GAGGCGGTCATGGGAGAGCTGATGTCGTGAGCCATGTGGCAGACGACTGATTTGGCACTAC 670
 Db 1332 CGCGCGCAAAAGGCTTCACAGCAATGATGCGCCAGTCCGCTCGCGCGGAGCAATTACAG 1273
 QY 671 AGCTTGGGGGGGCAATTTCCACGGGTG---AGCGGGCGCGGGTCTCCATCCGACGCCACTG 727
 Db 1272 CGCTACCGCGACATGTTCTTCGGGGGGCCAGCGCCAGCGCATCGCCTCGGCGCGGCATG 1213
 QY 728 CTCACAGGATCTTAAAGCTATGCTGTTTGATGAGCCAAACCAAGAGGCTCGACTCATGACT 787
 Db 1212 ATGTGCAACCAAGGTGCTGTGGCGGAGAGCGACGTCGGGCGCTCGACGTGTGATC 1153
 QY 788 GCTAATCATGATGTGTCGTCCTCCGCTGGTGAAGCTGGCTGCGAAGACGAATTTGGTTTC 847
 Db 1152 CAGGCCCAAGTACTGAACCTTTTATATGAGCTCGAGAGCAGATTCGCAACGCCCTACGTG 1093
 QY 848 ACCATTACCAAGCCCCGTTCTGAAGCTTTTACAGCTCTTTGACAAAATTTGCATCCAGAC 907
 Db 1092 TTTCATTCGCAACACTGGGGGTGTGTCGCGACAGTCGCGACGACGTCCTGTGTGATGTAC 1033
 QY 908 TTGCG 912
 Db 1032 CTCGG 1028

```

US-09-252-991A-13436/C
RESULT 10
US-09-252-991A-13436/C
Sequence 13436, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13436
LENGTH: 1509
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13436

```

Query Match	2.48;	Score 55.6;	DB 4;	Length 1509
Best Local Similarity	46.08;	Pred. No. 2.5e-05;		

[illegible]

Db 1669 CGCAGAGCGGTATCGGCTACCTGCCGAGAGCGTGGCGCTGTTCCATGGCAGC----- 1722
QY 539 CGGAGAGCGTGCATACACCGGCGTGTGGCCATCGGCGGCGCAATCCCGCTCTTC 598
Db 1723 -----CTGCGGAGACAACCTCAACCTGAGAAAGCGCCCTTGCGGAGAGAAATGCTG 1776
QY 599 CAGAAAGAGTGAAGCGCTCATATGAGAGAGCTAGTCTGAGCCATGTGGCAGACGACTG 1776
Db 1777 GAGACCTGAGCGGCTCGCGCCGCTGCTGCTGCGGCGCAACCGCGCTGGGCGCTGGAC 1836
QY 659 ATTGGCACTATACAGTGGGGGAGTTCACAGGCGGAGCGCGCGGCTGTCATGCA 718
Db 1837 ATCCGATCAGAGGAGGAGCGGCGGCTGCTGCGGCGCAACCGCGCTGGGCGCTGGCC 1896
QY 719 GCCAGCTGCTCCAGATCTTAAGGTCATGTTTGTATGAGCAACGACGAGCGCTGGAC 778
Db 1897 CGGTCCTCTACAGAGACCTCGATCTGCTGCTGAGAGCGGCGGCGCTGTCGAC 1956
QY 779 TGCATGACTGTAATCAGATTGCTGCTGCTGCTGAGAACTGCTGAGAGAAAGCAATT 838
Db 1957 CAGGCGAGGAGAAAGAGTACGACTGACAGCAATGTTGGGCAAGCGGCAACCTG 2016
QY 839 GTGTTTCAC 850
Db 2017 GTCATCACACC 2028

RESULT 13

US-09-252-991A-13575
; Sequence 13575, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13575
; LENGTH: 2292
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13575

Query Match 2.3%; Score 54; DB 4; Length 2292;
Best Local Similarity 45.3%; Pred. No. 8.6e-05;
Matches 250; Conservative 0; Mismatches 290; Indels 12; Gaps 1;

QY 299 ACCAGGAGATCTCAAGATGTCCTGTACGTGAGAGCGGAGATCATGTGCATC 358
Db 1460 AGCCCGCGGTGTACAGCTGACGAGCGCTCAACATCCGCGGAGCGGCGGTGGCGCTG 1519
QY 359 CTAGGAAGCTCAGGCTCCGGGAAACACGCTCTGAGCGCAATGTCGGGAGGCGGG 418
Db 1520 CTGGGCGGCAACGCGCGGCAAGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1579
QY 419 CGCGGCGGCACTTCTGCGGAGGTGTGTGAAGGGCGGCGGTGGCGGAGAGAG 478
Db 1580 GCGCAGCGGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1639
QY 479 TTCAGAGAGCTCTTCTCTAGCTCTGAGAGAGAGACCTCTGAGAGAGCTCAACG 538
Db 1640 CGCAGCGGAGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1693
QY 539 CGCAGAGCGTGCATACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598
Db 1694 -----CTGCGGAGACAACCTCAACCTGAGAAAGCGGCGGCTGGGAGAGAACTGCTG 1747

QY 599 CAGAAAGAGTGAAGCGCTCATATGAGAGAGCTAGTGTGAGCCATGTGGCAGACGACTG 658
Db 1748 GAGACCTGAGAGCGGCTGCGGCTGCGGCGCTTGTCCGCGGCGCACCCGCTGGGCTGAGC 1807
QY 659 ATTGGCACTATACAGTGGGGGAGCTATTCACAGGCTGAGCGGCGGCTGTCATGCA 718
Db 1808 ATCCGATCAGAGGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1867
QY 719 GCCAGCTGCTCCAGATCTTAAGGTCATGTTTGTATGAGCAACGACGAGCGCTGGAC 778
Db 1868 CGGTCCTCTACAGAGACCTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1927
QY 779 TGCATGACTGTAATCAGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 838
Db 1928 CAGGCGAGGAGAAAGAGTACGACTGACAGCAATGTTGGGCAAGCGGCAACCTG 1987
QY 839 GTGTTTCAC 850
Db 1988 GTCATCACACC 1999

RESULT 14

US-09-252-991A-10998
; Sequence 10998, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10998
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10998

Query Match 2.3%; Score 53.4; DB 4; Length 993;
Best Local Similarity 48.3%; Pred. No. 8e-05;
Matches 249; Conservative 0; Mismatches 251; Indels 15; Gaps 3;

QY 306 AGATCTCAAGATGTCCTGTGTACGTGAGAGCGGAGATCATGTGCATCTAGGA 365
Db 50 AGCCCTGAGCAACATCAACCTGATATCCAGAGCGGCAACTGTGCGCTCTGCGCC 109
QY 366 GCTCAGGCTCGGGAACACACCTGCTGAGAGCGCAATGTCGGGAGGCTGGCGCGG 425
Db 110 CTTCCGCGGTGCGGAAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 166
QY 426 GGAATCTTCTGCGGAGGTGTGTGTGAAGCGGCGGCGCTGCGCGCGGAGCACTTCAG 485
Db 167 AGCGCGGCAACATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 223
QY 486 ACTGCTTCTCTAGCTCTGAGAGAGAGACCTGCTGAGAGAGCCATGAGAGAGCGG 545
Db 224 GCAACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 283
QY 546 CGTGTGACTACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 605
Db 284 AGTGTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
QY 606 AGTGTGAGCGCTCATGAGAGAGAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 665
Db 344 AGCGCAAGTCTCAGAGAGCTG-----CTCAACATGTGTGAGTGTGAGTGTGAGTGTG 394
QY 666 ACTTACAGCTTGGGGGAGCTTTCACGAGTGAAGCGGCGGCTGCTGCTGCTGCTGCTG 725

Db 395 ACCGCTATCCAGACCACTCTCGGCGGCGACGCGCATATCGCCCTGGCCGCGCGC 454
QY 726 TGCATCAGATCTTAAGTATCTGTTGATGAGCCACACAGCCCTGGACATCATGA 785
Db 455 TGGCGGTGAGCCGAAATCTCTCTCGACGAACTTCGCGCCCTCGACGCGCAAG 514
QY 786 CTGCTAATCAGATTGCTCTCTCTGATGAGTGA 820
Db 515 TACGCAAGAGACTCGCGCGCTGCTGCGCGCGCTG 549

RESULT 15

US-09-252-991A-10934
: Sequence 10934, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 3142
: SEQ ID NO 10934
: LENGTH: 1335
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10934

Query Match 2.3%; Score 53.4; DB 4; Length 1335;
Best Local Similarity 48.3%; Pred. No. 9,4e-05;
Matches 249; Conservative 0; Mismatches 251; Indels 15; Gaps 3;

QY 306 AGATCTCAAGATGTCCTTGTACGTGAGAGCGGCGCATCATGTCATCTAGAA 365
Db 27 AGGCCCTGGACATCATCTGATATCCAGAGCGGCACTGTGCGCCCTGCTCGGC 86
QY 366 GCTCAGGCTCCGGGAAACAGCGTGTGAGACCGCGCTGCGCGGAGCATGTCGGAGCGCGCGG 425
Db 87 CTTCCGGGTGCGAAAGACCACTGCTGCGATCATGCGCG--GGCTGAAACCCCGG 143
QY 426 GGAACCTTCTGCGGAGGTATGTGAAACGCGCGCGCTGCGCGGAGCATGTCAGG 485
Db 144 ACGCGGCGACATGCTGTCATGCGGAGAGACGTCGTCGACAGACGAGCTGC--GCGACC 200
QY 486 ACTGCTTCTCTACGTCTCTGACAGGACACCTGCTGAGCAGGCTTCACCGTGGCGAGA 545
Db 201 GCAACGTGCGCTTCTGTCACACATGACGCGCTGTCGCGCACATGACGCTGTCGACA 260
QY 546 CGCTGACATACCGCGCTGCGCATCCGCGCGCATCCCGGCTCTCTCCAGAGA 605
Db 261 ACGTCCCTTCTGCGCTGCGCATGAGCCCAAGGCGCGCGCGGAGTCCGCGATCA 320
QY 606 AGGTGAGCGCGTATGAGCAGAGCTGAGCCATGTCGACAGCCGACTGATGGA 665
Db 321 AGGCCAAGTTCAGAGCTG-----CTCAACATGTCGAGCTGACTGGCTCGCCG 371
QY 666 ACTACAGCTGGGGGCGATTCACAGGCTGAGCGCGCGGCTCTCCATCGACGCCAGC 725
Db 372 ACCGCTATCCAGCACTCTCGGCGCGCGAGCGGAGCTATGCGCTGCGCGCGCG 431
QY 726 TGCATCAGATCTTAAGTATCTGTTGATGAGCCACACAGGCGCTGAGCTCATGA 785
Db 432 TGGCGGTGAGCCGAAATCTCTCTCGACGAACTTCGCGCGCTCGACGCGCAAG 491
QY 786 CTGCTAATCAGATTGCTCTCTCTGATGAGTGA 820
Db 492 TACGCAAGAGCTGCGCGCGCTGCTGCGCGCGCTG 526

Search completed: July 27, 2003, 12:45:40
Job time: 165.622 secs